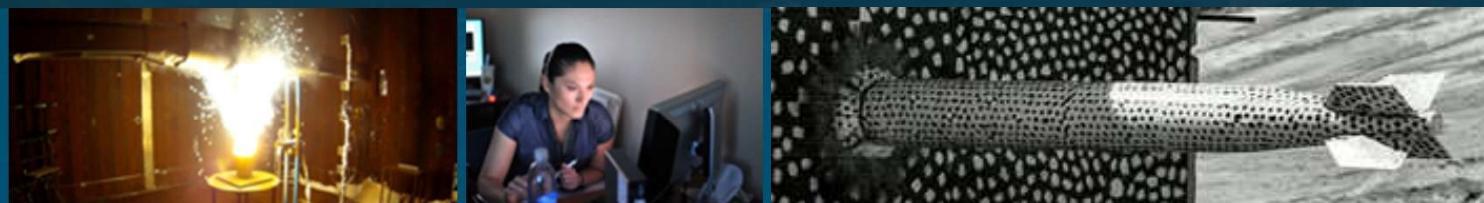


2022 American Physical Society (APS) March Meeting
Invited Session T42: Thermodynamics of Quantum Information Processing
Chicago, Illinois, USA



Exploring Fundamental Dissipation Limits of Reversible Computing Technologies from Non-equilibrium Quantum Thermodynamics



Thursday, March 17th, 2022

Michael P. Frank, Center for Computing Research

with Karpur Shukla, Lab for Emerging Technologies, Brown University

Approved for public release, SAND2022-3094 C



Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

(hide during talk)

Exploring Fundamental Limits of Reversible Computing Technologies from Nonequilibrium Quantum Thermodynamics

Thu. March 17, 12:06 p.m. – 12:42 p.m. CDT

Presenter: Mike Frank (Sandia National Laboratory)

Authors: Mike Frank (Sandia National Laboratory), Karpur Shukla (Brown University)

Landauer and Bennett famously argued that traditional *non-reversible* computational architectures suffer from a fundamental minimum energy dissipation (and entropy generation) that is required to carry out ordinary *logically irreversible* computational operations, but that alternative *reversible* computational architectures can circumvent this limit.

Over the years, questions have been raised regarding whether these observations remain valid when treated in a rigorous non-equilibrium thermodynamic framework. In recent work, we found that these classic statements do indeed remain valid for practical architectures when the role of correlations is properly taken into account. In particular, we have found that Müller's generalized framework of catalytic thermal operations provides a rigorous basis for these statements in a non-equilibrium context.

However, an important question remains regarding what fundamental limits on entropy generation can be shown to apply even to reversible computations. We conjecture that technology-independent limits on entropy generation in reversible computations (classical and quantum) can be formulated as a function of a number of relevant physical timescales, and outline our research plan for deriving these limits. As a first step, we discuss how to represent classical reversible operations in terms of a Lindbladian superoperator dynamics in a quantum Markovian framework.



Contributors to the larger effort

- Full group at Sandia:

- Michael Frank (Cognitive & Emerging Computing)
- Robert Brocato (RF MicroSystems)
- David Henry (MESA Hetero-Integration)
- Rupert Lewis (Quantum Phenomena)
- Nancy Missert (Nanoscale Sciences) – now retired
 - Matt Wolak (now at Northrop-Grumman)
- Brian Tierney (Rad Hard CMOS Technology)

Thanks are due to Sandia's LDRD program, DOE's ASC program, and the DoD/ARO ACI (Advanced Computing Initiative) for their support of this line of research!

Thanks are also due to the following colleagues & external collaborators:

- Karpur Shukla (CMU/Brown U.) *Key contributor to this work*
 - w. Prof. Jingming "Jimmy" Xu
- Hannah Earley (Cambridge U.)
- Erik DeBenedictis
- Joseph Friedman (UT Dallas)
 - with X. Hu, B.W. Walker, F. Garcia-Sanchez, P. Zhou, J.A.C. Incorviaz, A. Paller
- Kevin Osborn (LPS/JQI)
 - Liuqi Yu
- Steve Kaplan
- Rudro Biswas (Purdue)
 - Dewan Woods & Rishabh Khare
- Tom Conte (Georgia Tech/CRNCH)
 - Anirudh Jain, Gibran Essa
- David Guéry-Odelin (Toulouse U.)
- FAMU-FSU College of Engineering:
 - Sastry Pamidi (ECE Chair) & Jerris Hooker (Instructor)
 - 2019-20 students:
 - Frank Allen, Oscar L. Corces, James Hardy, Fadi Matloob
 - 2020-21 students:
 - Marshal Nachreiner, Samuel Perlman, Donovan Sharp, Jesus Sosa

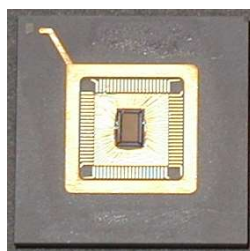
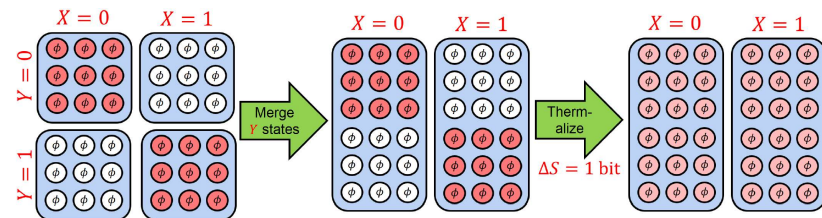
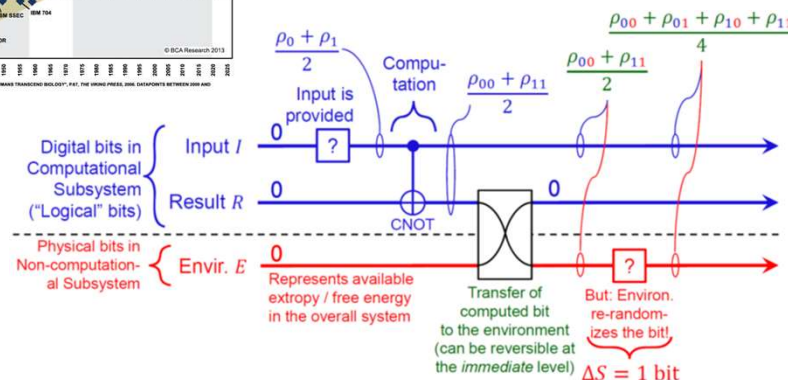
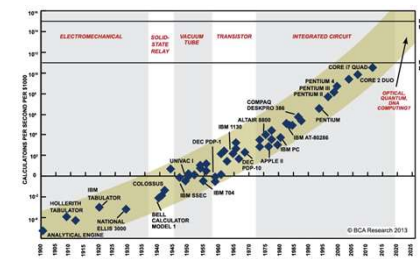


Talk Abstract/Outline

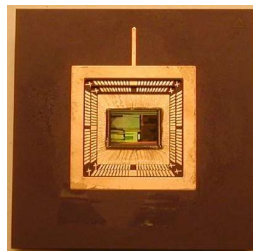


Exploring Fundamental Dissipation Limits of Reversible Computing Technologies from Non-equilibrium Quantum Thermodynamics

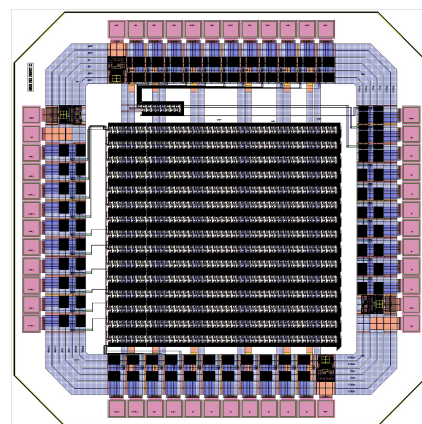
- Some classic results in the thermodynamics of computing:
 - Landauer (1961) – *Information loss* in computation implies entropy increase / energy dissipation.
 - Bennett (1973) – Information loss in computation can be *avoided* through *reversible computing*.
- Are these statements still valid even in a non-equilibrium thermodynamic context?
 - YES!** – We review why this is *rigorous* in a modern quantum thermodynamic theoretical context.
 - We also review a couple of types of engineering implementations already under development.
- Important open question:** How to formulate *general* limits on dissipation that apply even to reversible computations?
 - E.g.*, limits as a function of speed, temperature, coupling to environment, *etc.*
 - We outline an approach towards answering this question, and show some first steps.



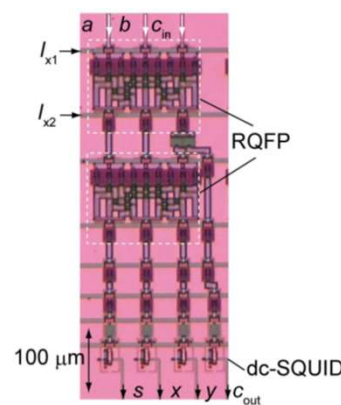
FlatTop
(MIT '96)



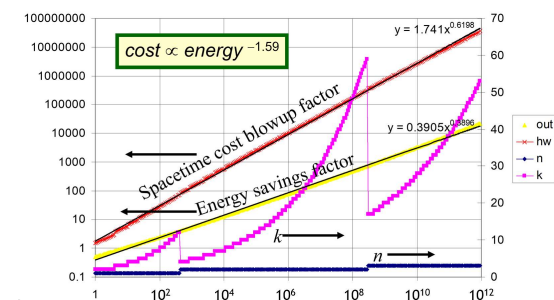
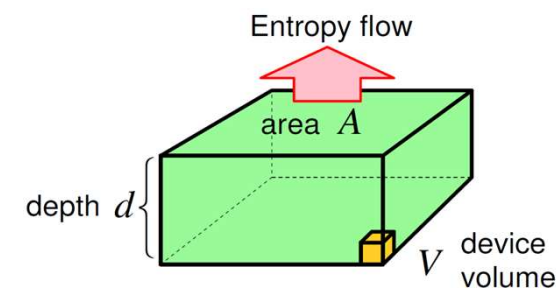
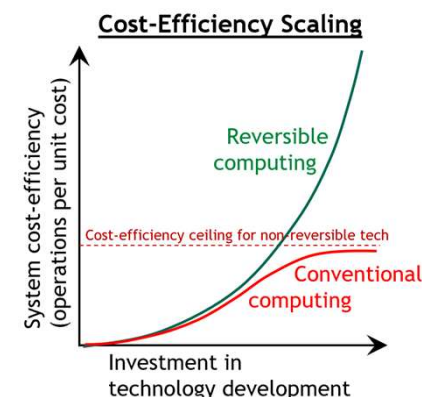
Pendulum
(MIT '99)



2LAL Shift Register (Sandia '20)



RQFP Full Adder (Yokohama '18)





I. Classic Thermodynamics of Computing – A Brief Review

Exploring Fundamental Dissipation Limits of
Reversible Computing Technologies from Non-
equilibrium Quantum Thermodynamics

Fundamental Physics of Computing—Earliest Roots (slide 1 of 2)



This topic can be placed on a firm theoretical foundation using tools from the field of *non-equilibrium quantum thermodynamics* (NEQT), the theoretical formulation of which derives from the mathematical foundations first laid down by von Neumann (1927). →

- However, even before von Neumann, the roots of modern stat. mech., thermodynamics and quantum theory were *already inseparable*.
- What we know today as “Boltzmann’s constant” k was actually first derived by Planck, in the *very same* analysis that simultaneously first resolved the value of what we now call “Planck’s constant” h .
- Statistical mechanics could never possibly have become a complete, coherent foundation for thermodynamics without the concomitant discovery of quantum mechanics! *Quantization is crucial*.

Some key foundational principles of NEQT are the following:

- *Unitary time evolution* of all closed systems (including the whole universe \mathcal{U})
- **NOTE:** von Neumann entropy $S = -\text{Tr}(\rho \ln \rho)$ is conserved by unitary transforms.
- Environment \mathcal{E} of an open system \mathcal{S} is treated as *independent* and *thermal*.
- Entropy *increase* can be viewed as merely a natural consequence of our inability as modelers to track quantum correlations (incl. entanglement) with (or within) any complex thermal environment \mathcal{E} .

Perspective is summarized in the definition of *thermal operations* derived from the (1955) *Stinespring Dilation Theorem*:

$$\rho_{\text{in},\mathcal{S}} \mapsto \mathbb{E}[\rho_{\text{in},\mathcal{S}}] := \text{Tr}_{\mathcal{E}}[\hat{U}_{t,\mathcal{S}\mathcal{E}}(\rho_{\text{in},\mathcal{S}} \otimes \tau_{\mathcal{E}})\hat{U}_{t,\mathcal{S}\mathcal{E}}^{\dagger}]$$

$$= \text{Tr}_{\mathcal{E}}[e^{-i\hat{H}t}(\rho_{\text{in},\mathcal{S}} \otimes \tau_{\mathcal{E}})e^{i\hat{H}t}].$$

Diagram annotations:

- Unitary evolution (curved arrow from $\hat{U}_{t,\mathcal{S}\mathcal{E}}$ to $\hat{U}_{t,\mathcal{S}\mathcal{E}}^{\dagger}$)
- Trace out correlations w. environment (arrow from $\text{Tr}_{\mathcal{E}}$ to the first $\text{Tr}_{\mathcal{E}}$)
- Thermal state of environment (arrow from $\tau_{\mathcal{E}}$ to the second $\text{Tr}_{\mathcal{E}}$)

Thermodynamik quantenmechanischer Gesamtheiten.

Von

J. v. Neumann, Berlin.

Vorgelegt von Max Born in der Sitzung vom 11. November 1927.

Einleitung.

I. In meiner Arbeit „Wahrscheinlichkeitstheoretischer Aufbau der Quantenmechanik“¹⁾ wurde gezeigt, wie die quantenmechanische Statistik aus einigen einfachen und rein qualitativen physikalischen Grundannahmen²⁾, sowie dem folgenden formalen Prinzip: die physikalischen Größen α eines gegebenen Systems \mathcal{S} entsprechen eindeutig und umkehrbar den (hermiteisch-)symmetri-

9. Ueber das Gesetz der Energieverteilung im Normalspectrum; von Max Planck.

(In anderer Form mitgeteilt in der Deutschen Physikalischen Gesellschaft, Sitzung vom 19. October und vom 14. December 1900, Verhandlungen 2. p. 202 und p. 237. 1900.)

Einleitung.

Die neueren Spectralmessungen von O. Lummer und E. Pringsheim¹⁾ und noch auffälliger diejenigen von H. Rubens und F. Kurlbaum²⁾, welche zugleich ein früher von H. Beckmann³⁾ erhaltenes Resultat bestätigten, haben

Hieraus und aus (14) ergeben sich die Werte der Naturkonstanten:

$$(15) \quad h = 6,55 \cdot 10^{-27} \text{ erg} \cdot \text{sec},$$

$$(16) \quad k = 1,346 \cdot 10^{-16} \frac{\text{erg}}{\text{grad}}.$$

Das sind dieselben Zahlen, welche ich in meiner früheren Mitteilung angegeben habe.

Fundamental Physics of Computing—Earliest Roots (slide 2 of 2)

“Shannon’s” 1948 entropy formula $H = -\sum p \log p$ was historically rooted in Boltzmann’s 1872 “H-theorem” (cf. E^* quantity below)

- Its general importance was *already* well established in statistical mechanics by the time of von Neumann’s (1920s) work on quantum thermodynamics.

However, Shannon did introduce some key new concepts such as mutual information, $I(X; Y) = H(X) + H(Y) - H(X, Y)$.

- The concept that information-bearing digital states can be identified with *sets* of (digitally interpreted) microstates also dates back to this era.

NOTE: Shannon *never once* addressed energy *dissipated*, only *invested*.

- There is nothing in Shannon’s (or von Neumann’s) work that contradicts RC.

Weitere Studien über das Wärmegleichgewicht unter Gas-molekülen.

Von Ludwig Boltzmann in Graz.

(Mit 4 Holzschnitten.)

(Vorgelegt in der Sitzung am 10. October 1872.)

Die mechanische Wärmetheorie setzt voraus, dass sich die Moleküle der Gase keineswegs in Ruhe, sondern in der lebhaftesten Bewegung befinden. Wenn daher auch der Körper seinen Zustand gar nicht verändert, so wird doch jedes einzelne seiner Moleküle seinen Bewegungszustand beständig verändern, und

$$E^* = N \iint f^* \log f^* ds d\sigma.$$

The Bell System Technical Journal

Vol. XXVII

July, 1948

No. 3

A Mathematical Theory of Communication

By C. E. SHANNON

INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist¹ and Hartley²

Quantities of the form $H = -\sum p_i \log p_i$ (the constant K merely amounts to a choice of a unit of measure) play a central role in information theory as measures of information, choice and uncertainty. The form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics³ where p_i is the probability of a system being in cell i of its phase space. H is then, for example, the **H in Boltzmann’s famous H theorem.** We shall call $H = -\sum p_i \log p_i$ the entropy of the set of probabilities

³ See, for example, R. C. Tolman, “Principles of Statistical Mechanics,” Oxford, Clarendon, 1938.

Communication in the Presence of Noise*

CLAUDE E. SHANNON†, MEMBER, IRE

I. INTRODUCTION

A GENERAL COMMUNICATIONS system is shown schematically in Fig. 1. It consists essentially of five elements.

1. *An information source.* The source selects one message from a set of possible messages to be transmitted to the receiving terminal. The message may be of various types; for example, a sequence of letters or numbers, as in telegraphy or teletype, or a continuous function of time $f(t)$, as in radio or telephony.

2. *The transmitter.* This operates on the message in some way and produces a signal suitable for transmission to the receiving point over the channel. In teleph-

Summary—A method is developed for representing any communication system geometrically. Messages and the corresponding signals are points in two “function spaces,” and the modulation process is a mapping of one space into the other. Using this representation, a number of results in communication theory are deduced concerning expansion and compression of bandwidth and the threshold effect. Formulas are found for the maximum rate of transmission of binary digits over a system when the signal is perturbed by various types of noise. Some of the properties of “ideal” systems which transmit at this maximum rate are discussed. The equivalent number of binary digits per second for certain information sources is calculated.

* Decimal classification: 621.38. Original manuscript received by the Institute, July 23, 1940. Presented, 1948 IRE National Convention, New York, N. Y., March 24, 1948; and IRE New York Section, New York, N. Y., November 12, 1947.

† Bell Telephone Laboratories, Murray Hill, N. J.

Landauer's Principle from Statistical Physics & Information Theory

(For further details, see [arXiv:1901.10327](https://arxiv.org/abs/1901.10327))

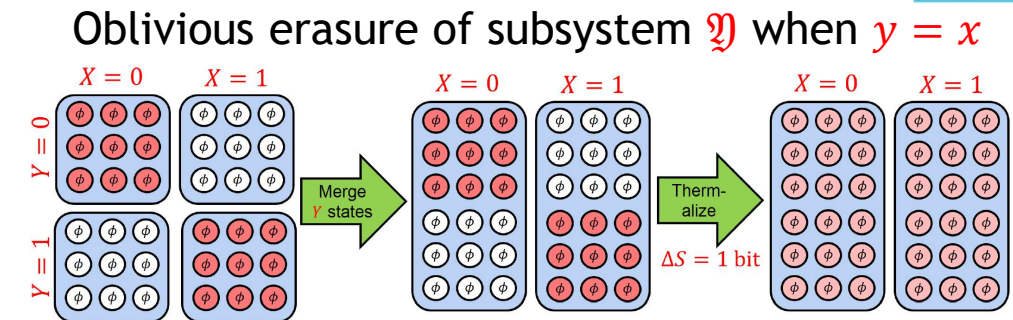


When stated **correctly**, proving Landauer's Principle is elementary...

- *I.e.*, it takes only a small handful of simple logical steps to prove;
- Depends **only** on basic facts of statistical physics and information theory.

Here's a **correct** statement of Landauer's Principle:

- Within any computational process composed out of **local, digital** primitive transformations, the **oblivious** (*i.e.*, isolated and unconditional) **erasure** (to a standard state) of a digital subsystem \mathfrak{Y} that possesses **marginal** digital entropy $H(Y)$ (entropy after restriction of the joint $\mathfrak{X}\mathfrak{Y}$ distribution to \mathfrak{Y}) and was **deterministically computed** from another subsystem \mathfrak{X} necessarily increases **total physical entropy** S by at least $H(Y)$. (We can also generalize from this a bit)
- **Corollary:** Free energy is reduced by $\Delta F = -H(Y) \cdot T$, and expulsion of entropy to environment results in heat emission $\Delta Q = H(Y) \cdot T$.
- **Generalization:** Any local reduction of \mathfrak{Y} 's marginal entropy by any amount $-\Delta H(Y)$ affects free energy and heat emission proportionately.



And here's a simple proof outline:

1. The Second Law of Thermodynamics ($\partial S / \partial t \geq 0$), together with the statistical definition of entropy, imply that microphysical dynamics *must* be **bijective** (this is reflected *e.g.* in the unitarity of quantum time-evolution).
2. Given that \mathfrak{Y} was computed **deterministically** from \mathfrak{X} , its conditional entropy $H(Y|X) = 0$, and therefore its marginal entropy is **entirely** accounted for by its mutual information with \mathfrak{X} , *i.e.*, $H(Y) = I(X; Y)$.
3. Because microphysics is bijective, local transformations **cannot destroy** the information $I(X; Y)$ but can only **eject** it out to some other subsystem (if not part of the machine's stable, digital state, it's in the thermal state).
4. Thermal environments, by definition, **don't preserve** correlation information at all (as reflected by, *e.g.*, thermal operations *a la* Stinespring); therefore, the total universe entropy gets increased by $\Delta S = I(X; Y) = H(Y)$.
 - This can be seen through the trace operation over \mathfrak{E} , or more simply by just observing that joint entropy $H(X, Y) = H(X) + H(Y) - I(X; Y)$ over two systems increases by $I(X; Y)$ if the original mutual information $I(X; Y)$ gets replaced with a new value $I'(X; Y) = 0$.

Basic Reversible Computing Theory

(For full proofs, see [arxiv:1806.10183](https://arxiv.org/abs/1806.10183))

Fundamental theorem of traditional reversible computing:

- A deterministic computational operation is (unconditionally) non-entropy-ejecting if and only if it is *unconditionally* logically reversible (i.e., injective over its entire domain).

Fundamental theorem of generalized reversible computing:

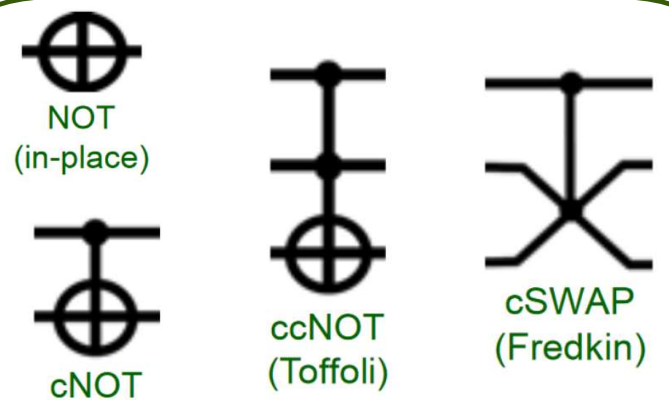
- A *specific* (contextualized) deterministic computational process is (specifically) non-entropy-ejecting if and only if it is *specifically* logically reversible (injective over the set of *nonzero-probability* initial states).
- Also, for any deterministic computational operation, which is conditionally reversible under some assumed precondition, then the entropy required to be ejected by that operation approaches 0 as the probability that the precondition is satisfied approaches 1.

Bottom line: To avoid requiring Landauer costs, it is *sufficient to just have logical reversibility if some specified preconditions are satisfied* (and then satisfy them).

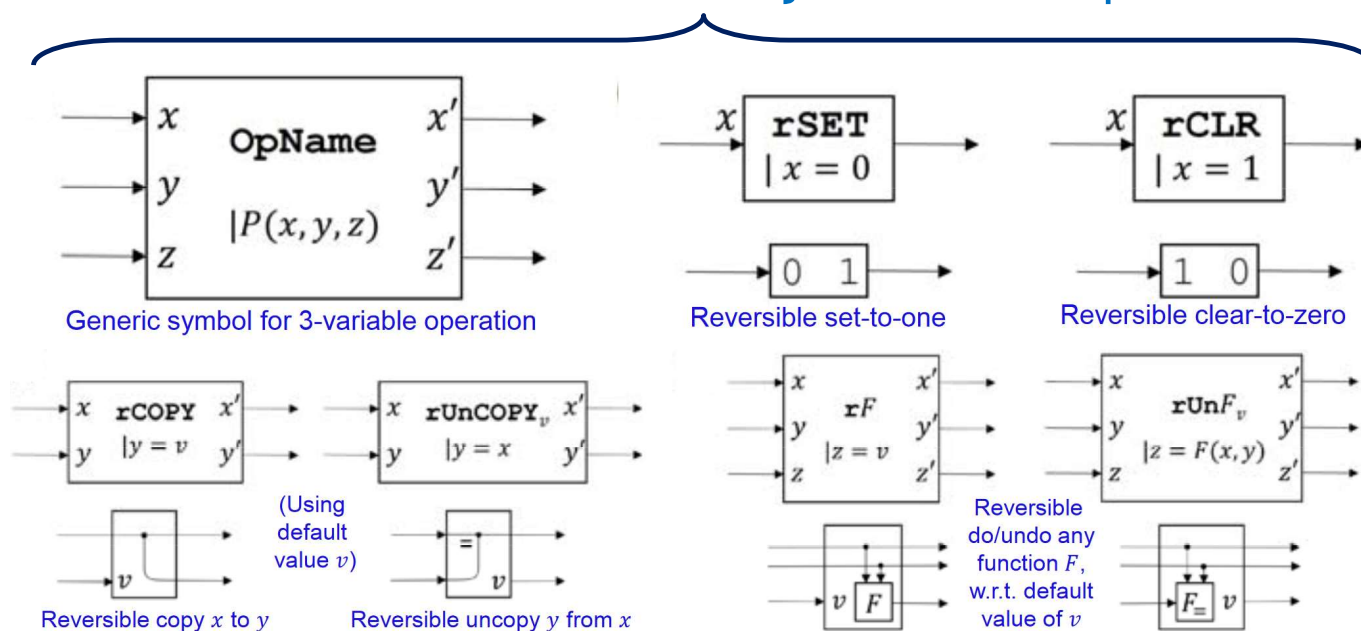
- This gives us a realistic (and more flexible!) basis for developing practical engineering implementations.
- An example of this is provided by fully adiabatic CMOS.



Traditional *Unconditionally* Reversible “Gates” (Operations)



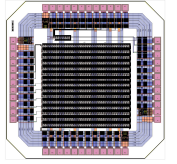
Some Generalized *Conditionally* Reversible Operations





II. Examples of Reversible Computing Technologies Currently Under Development

Exploring Fundamental Dissipation Limits of
Reversible Computing Technologies from Non-
equilibrium Quantum Thermodynamics



Perfectly Adiabatic Reversible Computing in CMOS

To approach ideal reversible computing in CMOS...

We must aggressively eliminate *all* sources of non-adiabatic dissipation, including:

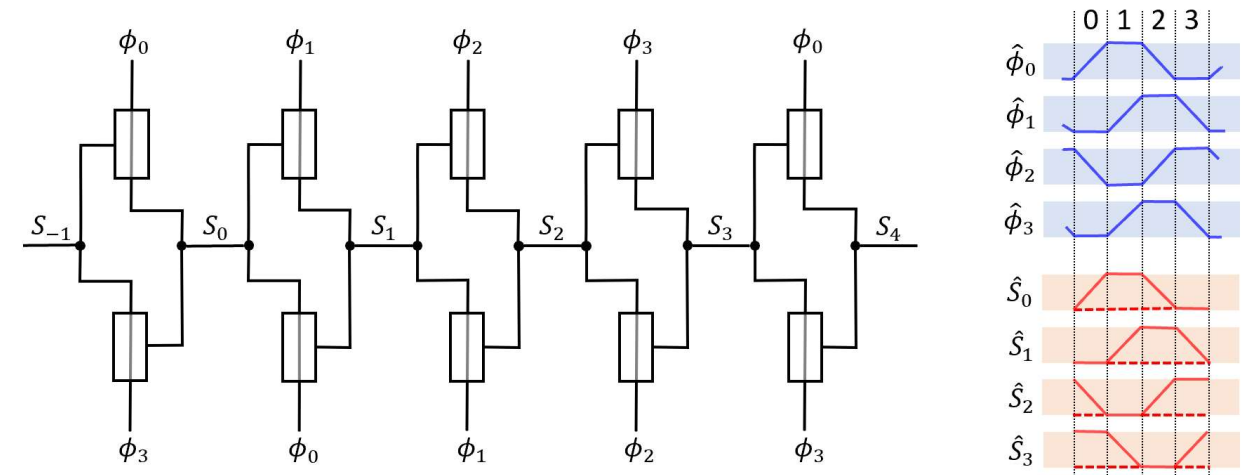
- Diodes in charging path, “sparking,” “squelching,”
 - Eliminated by “**truly, fully adiabatic**” design. (E.g., CRL, 2LAL).
 - Can suffice to get down to a few aJ (10s of eV) even *before* voltage optimization.
- Voltage level mismatches that dynamically arise on floating nodes before reconnection.
 - Eliminated by static, “**perfectly adiabatic**” design. (E.g., S2LAL).

We must also aggressively minimize standby power dissipation from leakage, including:

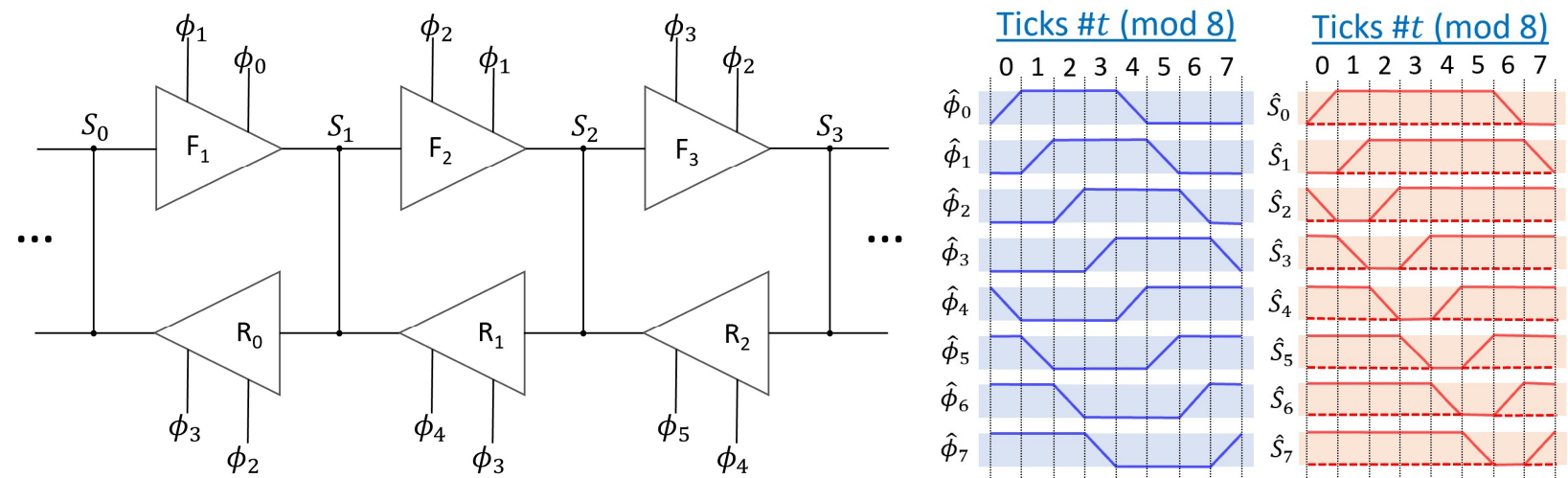
- Subthreshold channel currents.
 - Ultra-low- T (e.g. 4K) operation helps with this.
- Tunneling through gate oxide.
 - E.g., use thicker gate oxides.

Note: (Conditional) logical reversibility *follows from* perfect adiabaticity!

Shift Register Structure and Timing in 2LAL



Shift Register Structure and Timing in S2LAL



(arxiv:2009.00448)

Examples of S2LAL Logic Gates

14-transistor AND gate, 16-transistor OR gate.

- Carefully designed to ensure that each internal node is always connected to either a constant or variable source.
- The structures shown are minimal, given the design constraints.

Inverting gates are done easily, by using signal pairs for complementary symbols:

- $\text{NOT}(A^1) = \text{BUFFER}(A^0)$
- $\text{NAND}(A^1, B^1) = \text{OR}(A^0, B^0)$
- $\text{NOR}(A^1, B^1) = \text{AND}(A^0, B^0)$

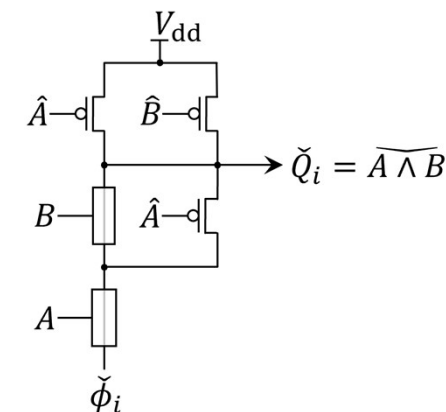
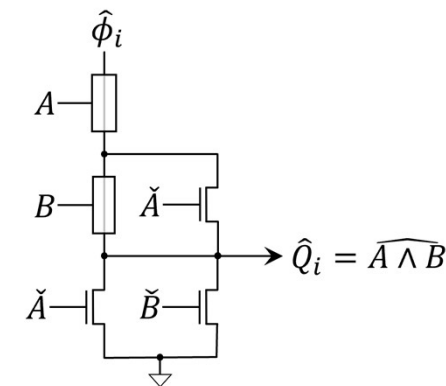
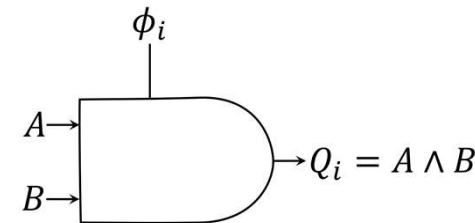
Also! Erik DeBenedictis invented an optimization to S2LAL that can compute the inverses as-needed, rather than always keeping both the 0,1 signal pairs around all the time:

- See <https://zettaflops.org/zf004/>.

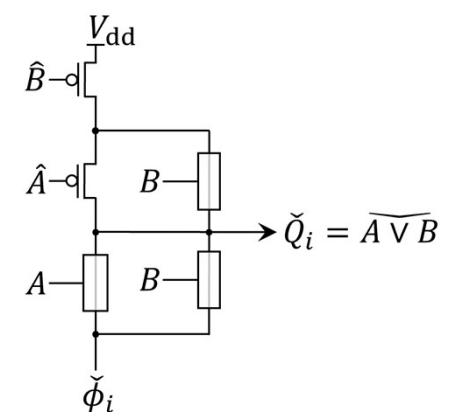
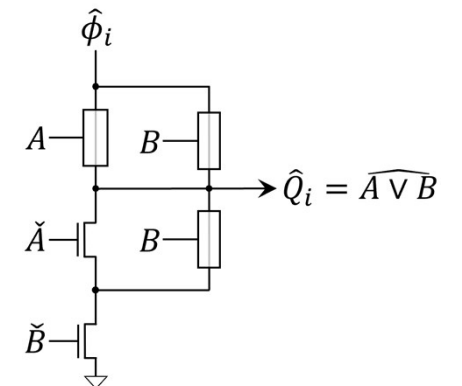
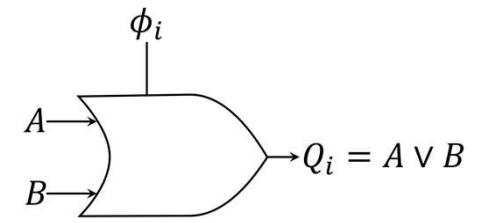
(These gates correctly support conditionally reversible operations.)



AND



OR



Minimum Energy Scaling for Adiabatic CMOS

From M. Frank & K. Shukla,
doi:10.3390/e23060701



Appendix A. Minimum-Energy Scaling for Classical Adiabatic Technologies

In this appendix, we briefly present the derivation for the scaling of minimum energy dissipation for reversible technologies such as RA-CMOS (Section 2.3.1) that obey classic adiabatic scaling and that can be characterized in terms of relaxation and equilibration timescales.⁴⁴

First, we assume (as is the case for “perfectly adiabatic” technologies such as [48]) that the total energy dissipation per clock cycle E_{diss} in a reversible circuit can be expressed as a sum of *switching losses* and *leakage losses*,

$$E_{\text{diss}} = E_{\text{sw}} + E_{\text{lk}}, \quad (\text{A1})$$

and further, that switching and leakage losses depend on the signal energy E_{sig} and transition time t_{tr} approximately as follows:

$$E_{\text{sw}} \simeq E_{\text{sig}} \cdot c_{\text{sw}} \cdot \frac{\tau_r}{t_{\text{tr}}}, \quad (\text{A2})$$

$$E_{\text{lk}} \simeq E_{\text{sig}} \cdot c_{\text{lk}} \cdot \frac{t_{\text{tr}}}{\tau_e}, \quad (\text{A3})$$

where τ_r, τ_e are the relaxation and equilibration timescales, respectively, and $c_{\text{sw}}, c_{\text{lk}}$ are small dimensionless constants characteristic of a particular reversible circuit in a specific family of technologies, such as [48]. In practice, although these specific formulas are only approximate, they approach exactness in the regime $\tau_r \ll t_{\text{tr}} \ll \tau_e$.

Then, now treating (A2), (A3) as exact, we can write:

$$E_{\text{diss}} = E_{\text{sig}} \left(c_{\text{sw}} \tau_r \cdot \frac{1}{t_{\text{tr}}} + \frac{c_{\text{lk}}}{\tau_e} \cdot t_{\text{tr}} \right). \quad (\text{A4})$$

We can collect the constants, absorbing them into adjusted timescales $\tau'_r = c_{\text{sw}} \tau_r$ and $\tau'_e = \tau_e / c_{\text{lk}}$, so

$$E_{\text{diss}} = E_{\text{sig}} \left(\tau'_r \cdot \frac{1}{t_{\text{tr}}} + \frac{1}{\tau'_e} \cdot t_{\text{tr}} \right). \quad (\text{A5})$$

Setting the derivative of (A5) with respect to t_{tr} equal to zero, we find that E_{diss} is minimized when

$$\tau'_r \frac{1}{t_{\text{tr}}^2} = \frac{1}{\tau'_e}, \quad (\text{A6})$$

Upshot for CMOS: As each device's leakage conductance I_{off} is decreased, the equilibration timescale τ_e increases, and the technology's minimum energy (given perfectly adiabatic, reversible designs) scales down with square-root proportionality.

$$E_{\text{diss,min}} \propto \frac{1}{\sqrt{\tau_e}} \propto \sqrt{I_{\text{off}}}$$

or in other words, when

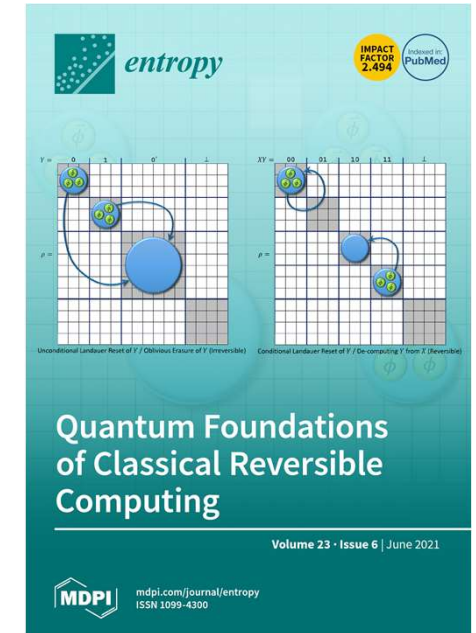
$$t_{\text{tr}} = \sqrt{\tau'_r \tau'_e}, \quad (\text{A7})$$

at which point E_{sw} and E_{lk} are equal. The minimum energy dissipation per cycle is then

$$E_{\text{diss}} = 2E_{\text{sig}} \sqrt{\frac{\tau'_r}{\tau'_e}}. \quad (\text{A8})$$

Thus, for any given reversible circuit design in a family of technologies with given values of the constants $c_{\text{sw}}, c_{\text{lk}}$, in order for E_{diss} to approach 0 as the technology develops, we must have that the ratio of equilibration/relaxation timescales $\tau_e / \tau_r \rightarrow \infty$, and, if the relaxation timescale τ_r is fixed, this implies that also the (minimum-energy) value of the transition time $t_{\text{tr}} \rightarrow \infty$. These requirements were mentioned in Section 2.3.1.

More specifically, in order to increase the peak energy efficiency of a reversible circuit by a factor of $N \times$, in a given family of technologies obeying classic adiabatic scaling, this requires that the timescale ratio τ_e / τ_r must be increased by $N^2 \times$, and (assuming τ_r is fixed) the transition time t_{tr} for minimum energy will increase by $N \times$.



Latest Results from the “Adiabatic Circuits Feasibility Study” Simulation Efforts at Sandia, funded via NSCI



Created schematic-level fully-adiabatic designs for Sandia’s in-house CMOS processes, including:

- Older, 350 nm process (**blue** curve)
 - FET widths = 800 nm
- Newer, 180 nm process (**orange, green** curves)
 - FET widths = 480 nm

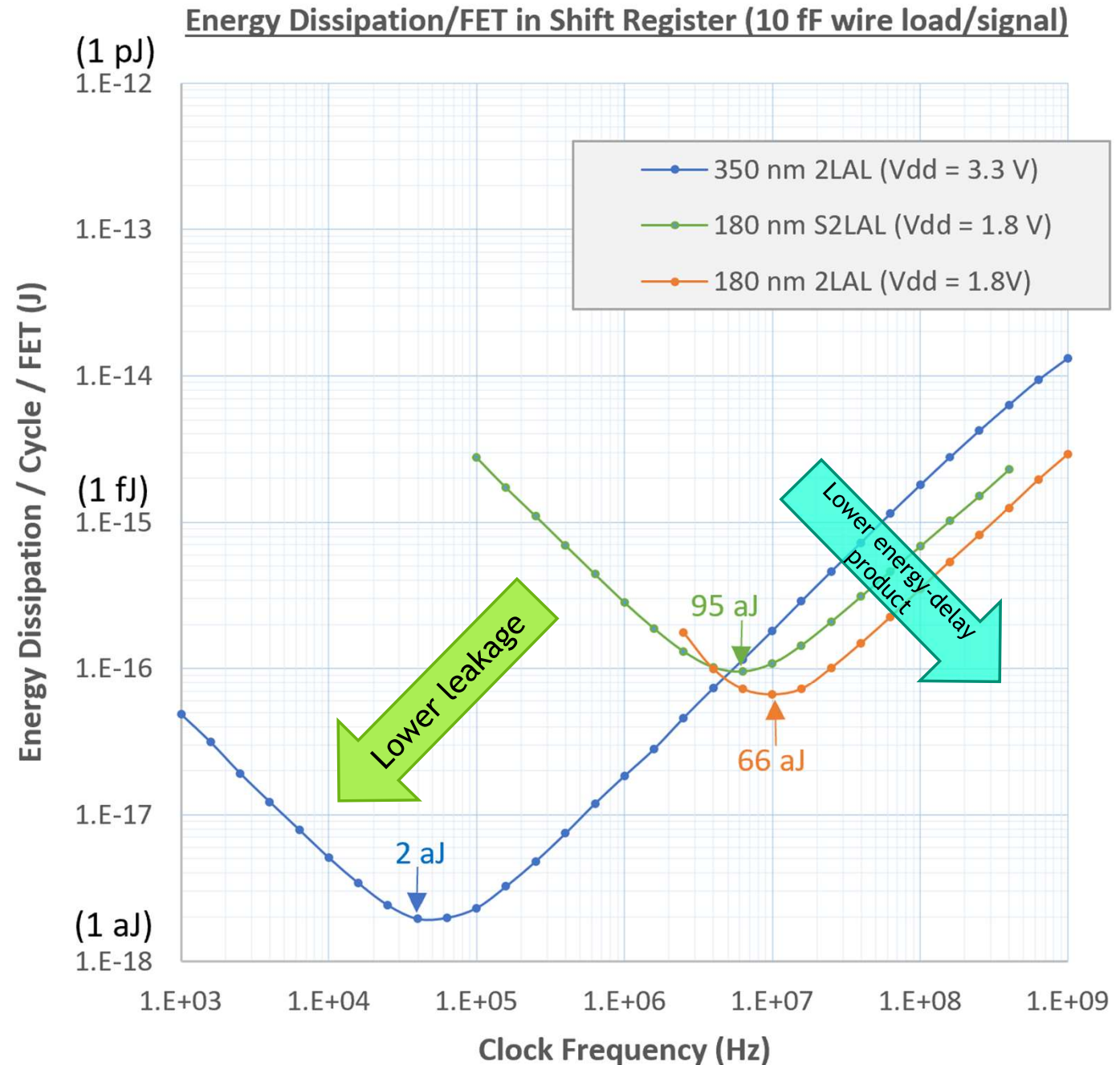
Plotted energy dissipation per-transistor in shift registers at 50% activity factor (alternating 0/1)

- 2LAL (**blue, orange** curves)
- S2LAL (**green** curve)

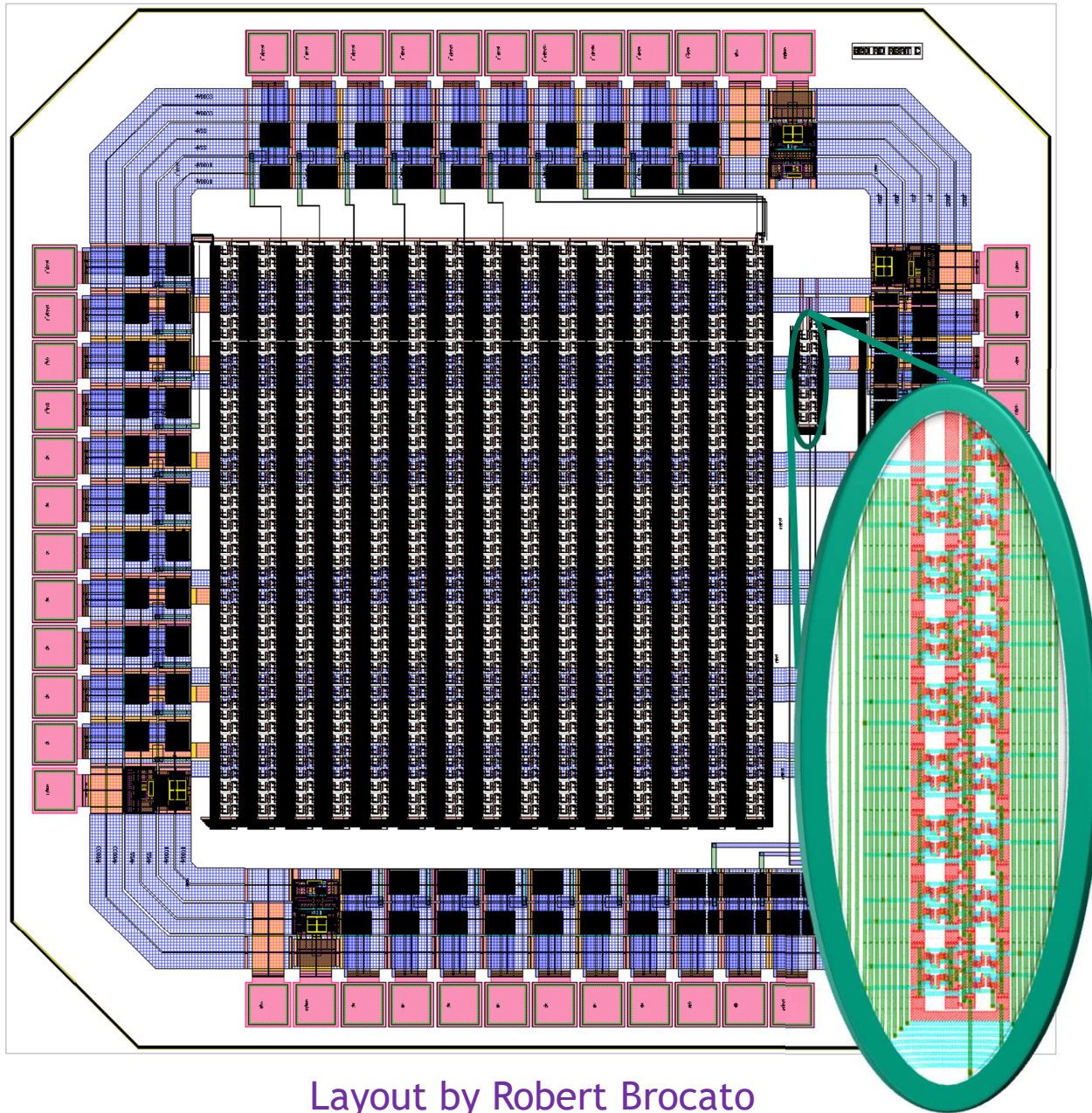
In all of these Cadence/Spectre simulations,

- We assumed a 10 fF parasitic wiring load capacitance on each interconnect node.
- Logic supply (V_{dd}) voltages were taken at the processes’ nominal values.
 - 3.3V for the 350nm process; 1.8V in the 180nm process.

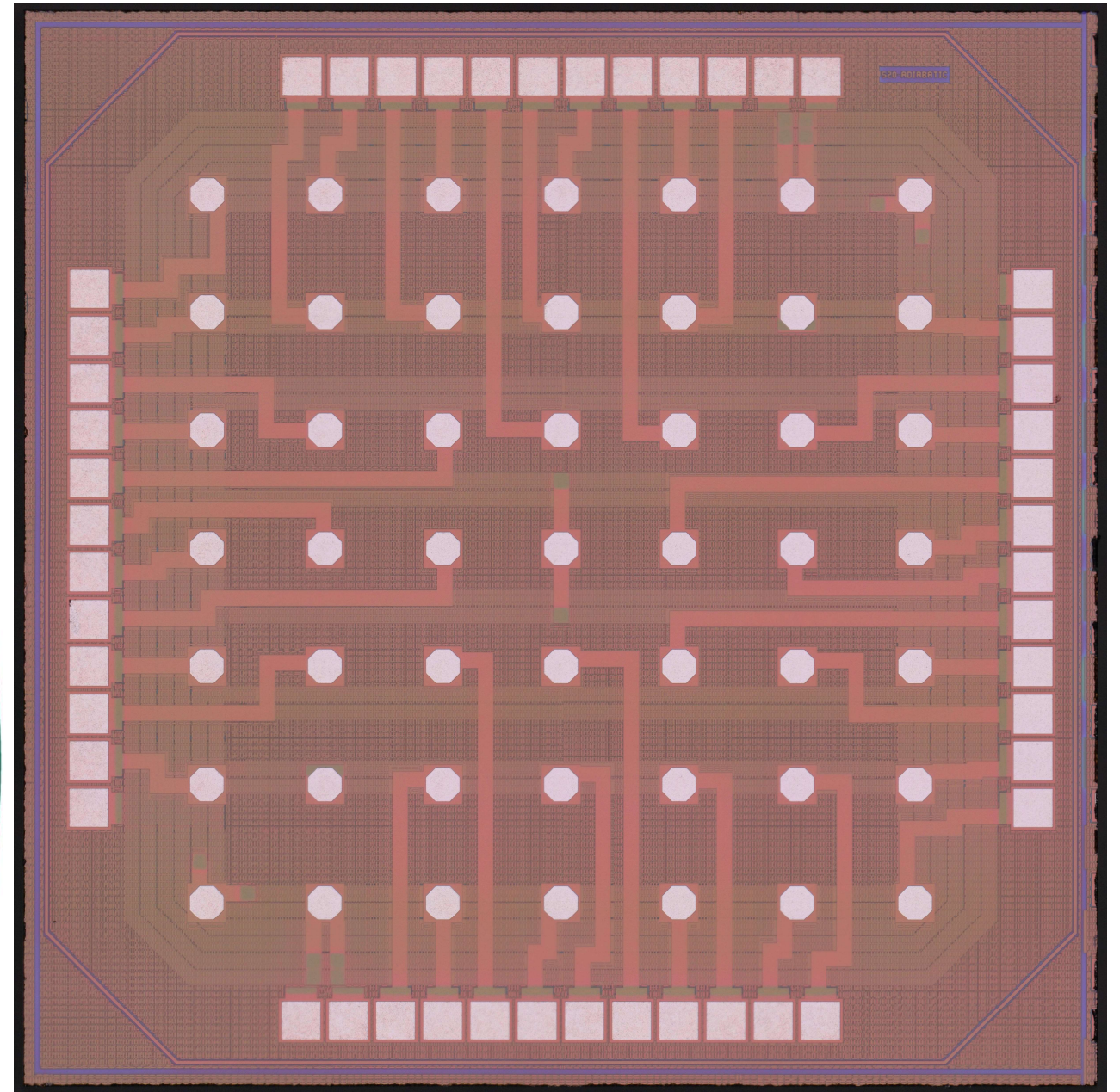
We expect these results could be significantly improved by exploring the parameter space over possible values of V_{dd} and V_{sb} (substrate bias).



2LAL Test Chip



Layout by Robert Brocato



Fabricated Die (photo credit: Darlene Udoni)

Adiabatic Reversible Computing in Superconducting Circuits



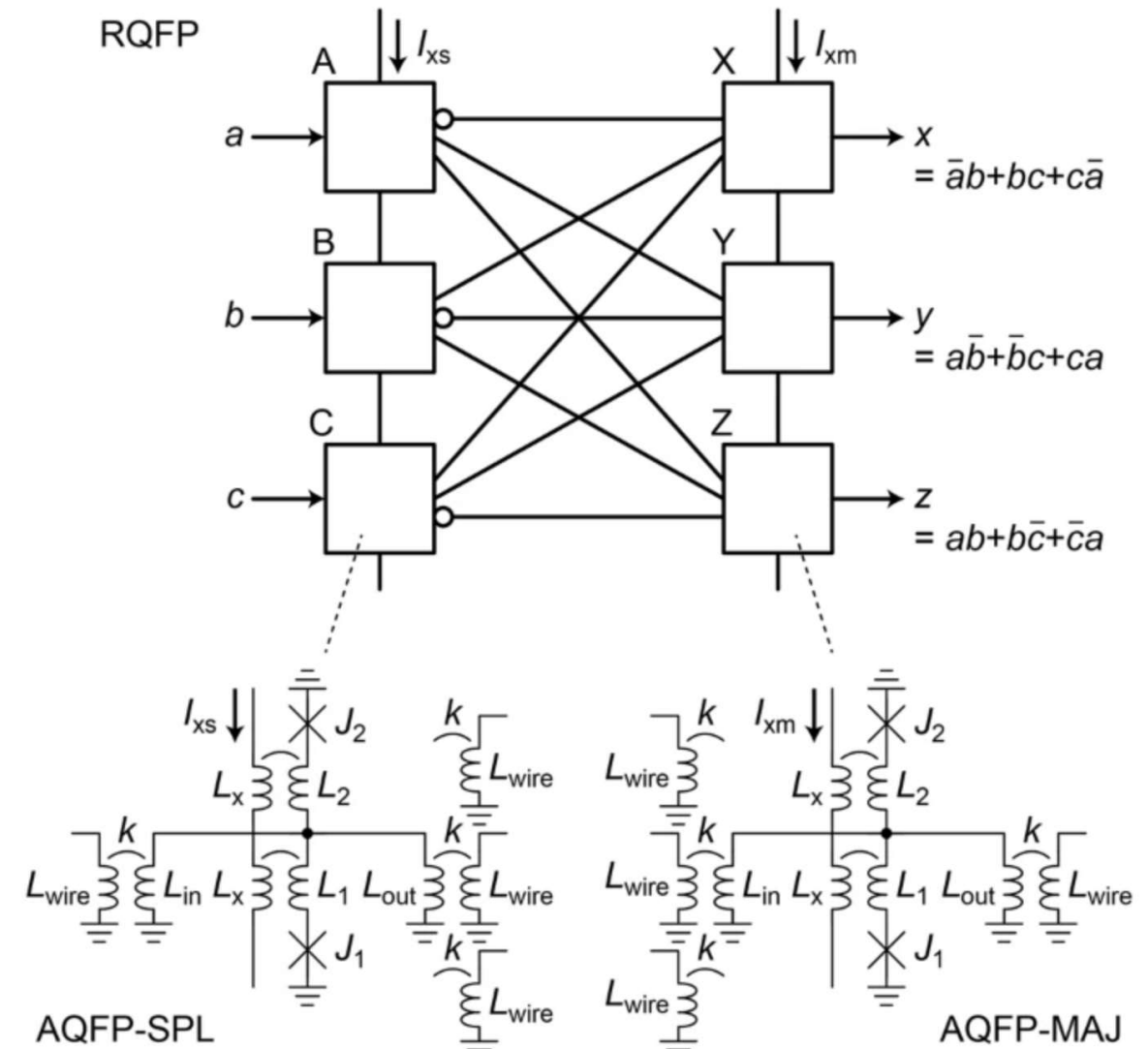
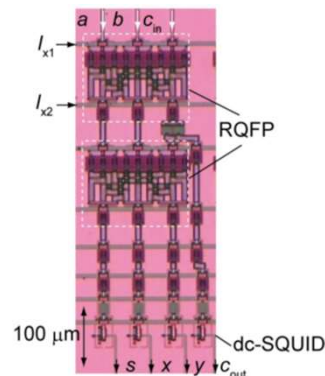
Work along this general line has roots that go all the way back to Likharev, 1977. (doi:10.1109/TMAG.1977.1059351)

- Most active group recently is Prof. Yoshikawa's group at Yokohama National University in Japan.

Logic style called *Reversible Quantum Flux Parametron* (RQFP).

- Shown at right is a 3-output *reversible majority gate*.
- Full adder circuits have also been built and tested.

Simulations indicate that RQFP circuits can dissipate $< kT \ln 2$ (even noting that $T = 4\text{K}$), at speeds on the order of 10 MHz



Existing Dissipation-Delay Products (DdP)— Adiabatic Reversible Superconducting Circuits

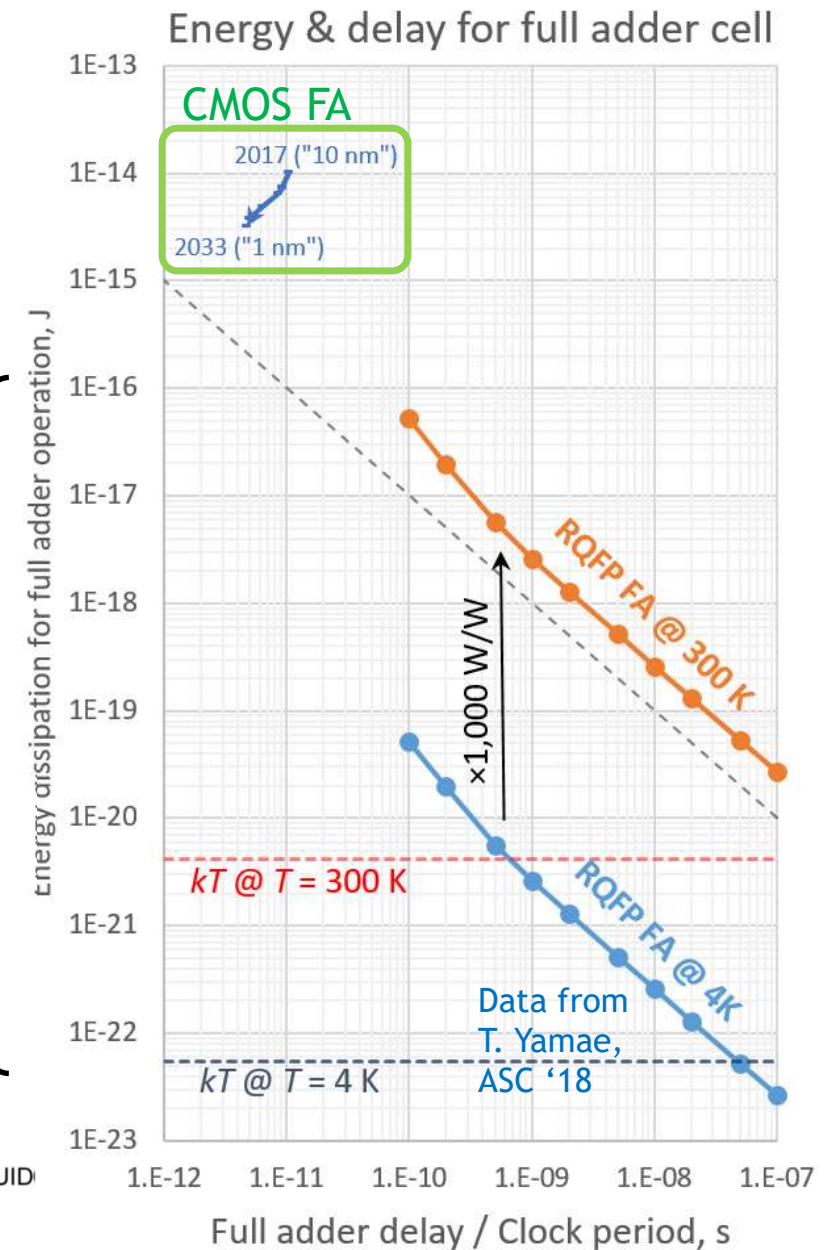
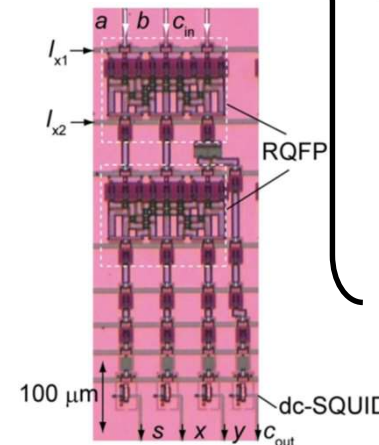
Reversible adiabatic superconductor logic:

- State-of-the-art is the **RQFP** (Reversible Quantum Flux Parametron) technology from Yokohama National University in Japan.
- Chips were fabricated, function validated.
- Circuit simulations predict DdP is $>1,000\times$ *lower* than even *end-of-roadmap* CMOS.
- Dissipation extends *far below* the 300K Landauer limit (and even below the Landauer limit at 4K).
- DdP is *still* better than CMOS even after adjusting by a conservative factor for large-scale cooling overhead ($1,000\times$).

Question: Could some *other* reversible technology do even better than this?

- We have a project at Sandia exploring one possible superconductor-based approach for this (more later)...
- But, what are the *fundamental* (technology-independent) limits, if any?

RQFP =
Reversible
Quantum Flux
Parametron
(Yokohama U.)



Ballistic Reversible Computing

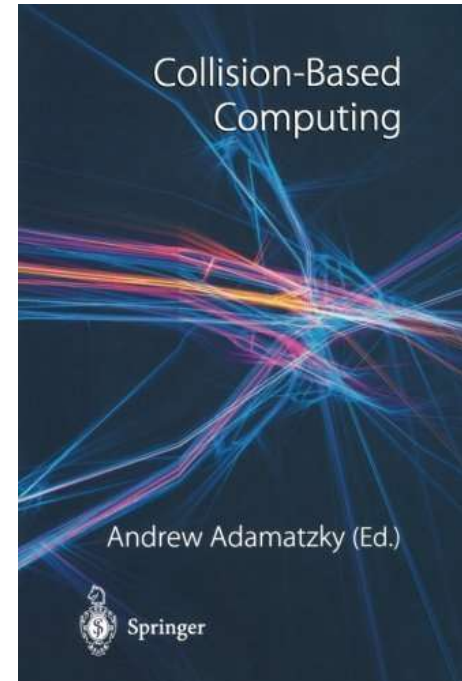
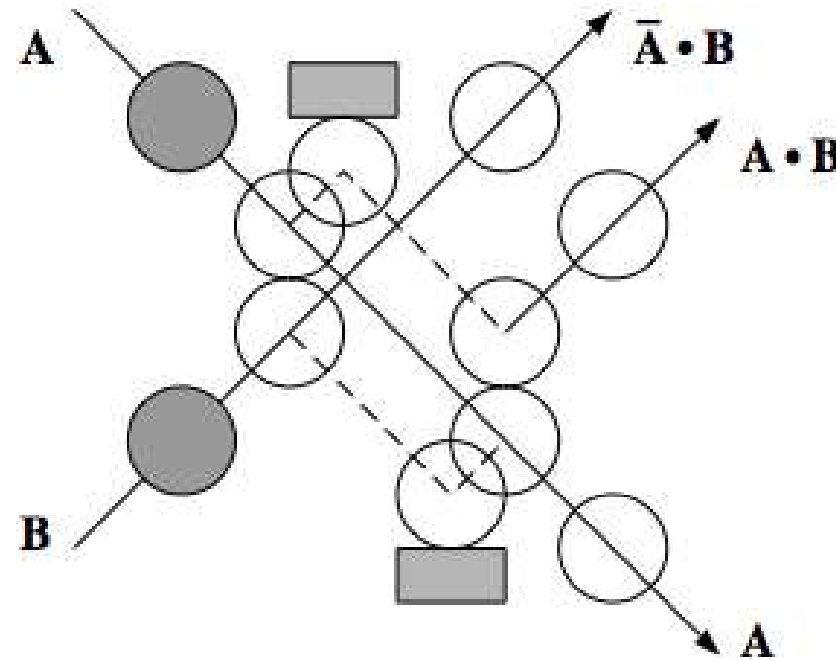
Can we envision reversible computing as a deterministic elastic interaction process?

Historical origin of this concept:

- Fredkin & Toffoli's *Billiard Ball Model* of computation ("Conservative Logic," IJTP 1982).
 - Based on elastic collisions between moving objects.
 - Spawned a subfield of "collision-based computing."
 - Using localized pulses/solitons in various media.

No power-clock driving signals needed!

- Devices operate when data signals arrive.
- The operation energy is carried by the signal itself.
 - Most of the signal energy is preserved in outgoing signals.



However, all (or almost all) of the existing design concepts for ballistic reversible computing invoke implicitly *synchronized* arrivals of ballistically-propagating signals...

- Making that approach work in reality presents some serious difficulties, however:
 - Unrealistic in practice to assume precise alignment of signal arrival times.
 - Thermal fluctuations & quantum uncertainty, at minimum, are always present.
 - Any relative timing uncertainty leads to chaotic dynamics when signals interact.
 - Exponentially-increasing uncertainties in the dynamical trajectory.
 - Deliberate *resynchronization* of signals whose timing relationship has become uncertain incurs an inevitable energy cost.

Can we come up with a *new* ballistic model of reversible computing that avoids these problems?

Ballistic Asynchronous Reversible Computing (BARC)



Problem: Conservative (dissipationless) dynamical systems tend to exhibit chaotic behavior...

- This results from direct nonlinear *interactions* between multiple continuous dynamical degrees of freedom (DOFs), which amplify uncertainties, exponentially compounding them over time...
- *E.g.*, positions/velocities of ballistically-propagating “balls”
 - Or more generally, any localized, cohesive, momentum-bearing entity: Particles, pulses, quasiparticles, solitons...

Core insight: To greatly reduce or eliminate this tendency towards dynamical chaos...

- Simply *avoid* direct interaction between continuous DOFs of different ballistically-propagating entities

Require localized pulses arrive *asynchronously*—and further, at clearly distinct, *non-overlapping* times

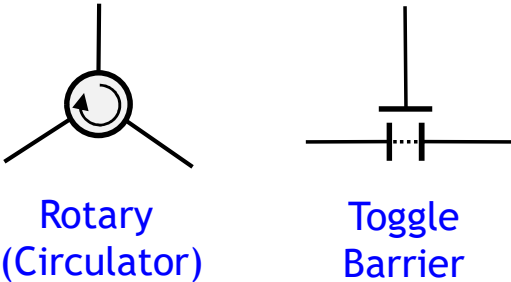
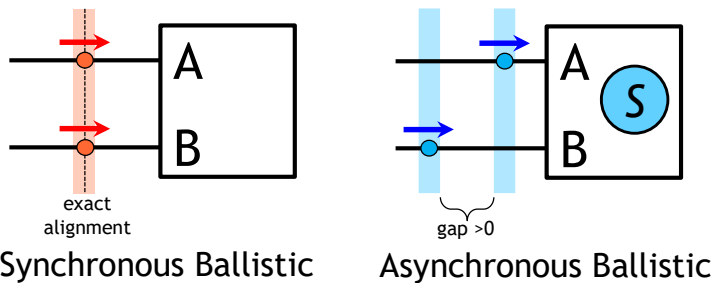
- Device’s dynamical trajectory is then *independent* of the precise (absolute *and* relative) pulse arrival times
 - As a result, timing uncertainty per logic stage can now accumulate only *linearly*, not exponentially!
 - Only relatively occasional re-synchronization will be needed
- For devices to still be capable of doing logic, they must now maintain an internal discrete (digitally-precise) state variable—a stable (or at least metastable) stationary state, *e.g.*, a ground state of a well

No power-clock signals, unlike in adiabatic designs!

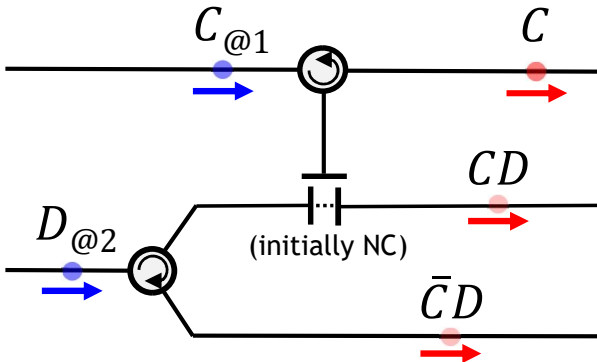
- Devices simply operate whenever data pulses arrive
- The operation energy is carried by the pulse itself
 - Most of the energy is preserved in outgoing pulses
 - Signal restoration can be carried out incrementally, or periodically

Goal of current effort at Sandia: Demonstrate BARC principles in an implementation based on fluxon dynamics in Superconducting Electronics (SCE)

(“BARCS”  effort)



Example BARC device functions



Example logic construction

Simplest Fluxon-Based (bipolarized) BARC Function



One of our early tasks: Characterize the simplest nontrivial BARC device functionalities, given a few simple design constraints applying to an SCE-based implementation, such as:

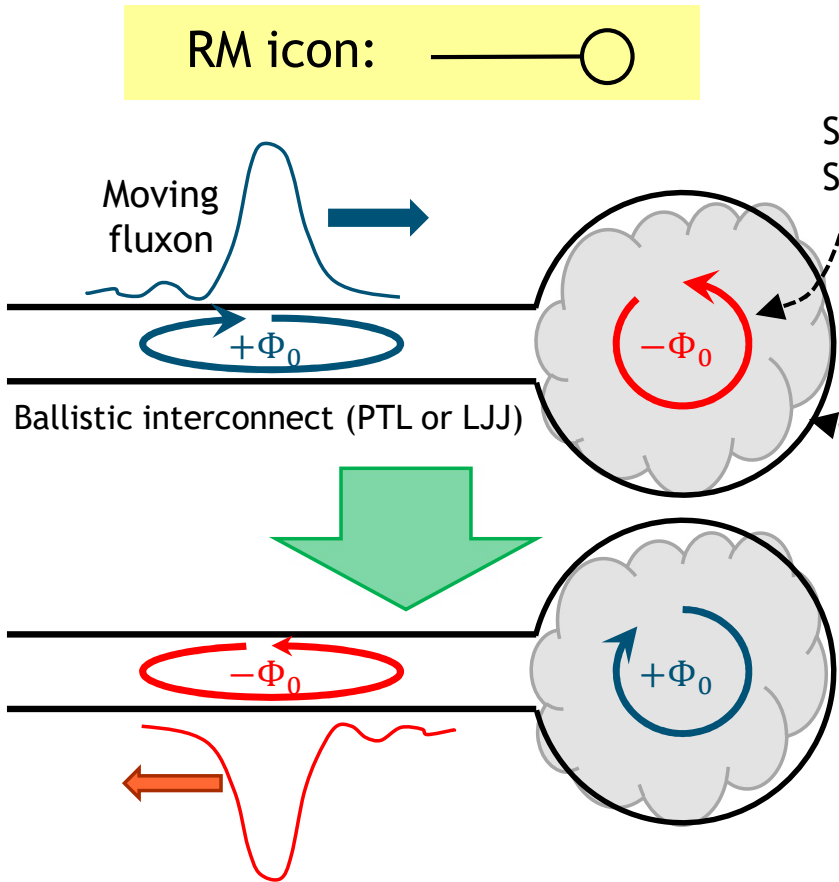
- (1) Bits encoded in fluxon polarity; (2) Bounded planar circuit conserving flux; (3) Physical symmetries.

Determined through theoretical hand-analysis that the simplest such function is the 1-Bit, 1-Port Reversible Memory (RM) Cell:

- Due to its simplicity, this was then the initial target for our subsequent detailed circuit design efforts...

RM Transition Table

Input Syndrome		Output Syndrome
+1(+1)	→	(+1)+1
+1(-1)	→	(+1)-1
-1(+1)	→	(-1)+1
-1(-1)	→	(-1)-1



- Some planar, unbiased, reactive SCE circuit w. a continuous superconducting boundary
- Only contains L's, M's, C's, and *unshunted* JJs
 - Junctions should mostly be *subcritical* (avoids R_N)
 - Conserves total flux, approximately nondissipative
- Desired circuit behavior (NOTE: conserves flux, respects T symmetry & logical reversibility):
- If polarities are opposite, they are swapped (shown)
 - If polarities are identical, input fluxon reflects back out with no change in polarity (not shown)
 - (*Deterministic*) *elastic 'scattering'* type interaction: Input fluxon's kinetic energy is (nearly) preserved in output fluxon

RM—First working (in simulation) implementation!



Erik DeBenedictis: “Try just strapping a JJ across that loop.”

- This actually works!

“Entrance” JJ sized to = about 5 LJJ unit cells ($\sim 1/2$ pulse width)

- I first tried it twice as large, & the fluxons annihilated instead...
 - “If a $15 \mu\text{A}$ JJ rotates by 2π , maybe $1/2$ that will rotate by 4π ” 🤔

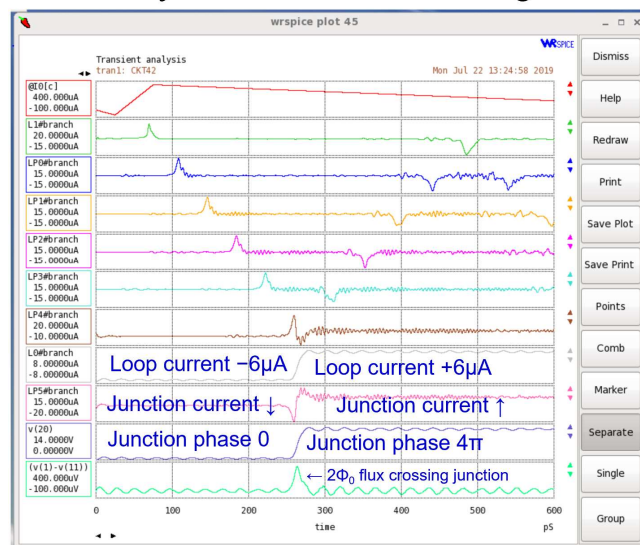
Loop inductor sized so ± 1 SFQ will fit in the loop (but not ± 2)

- JJ is sitting a bit below critical with ± 1

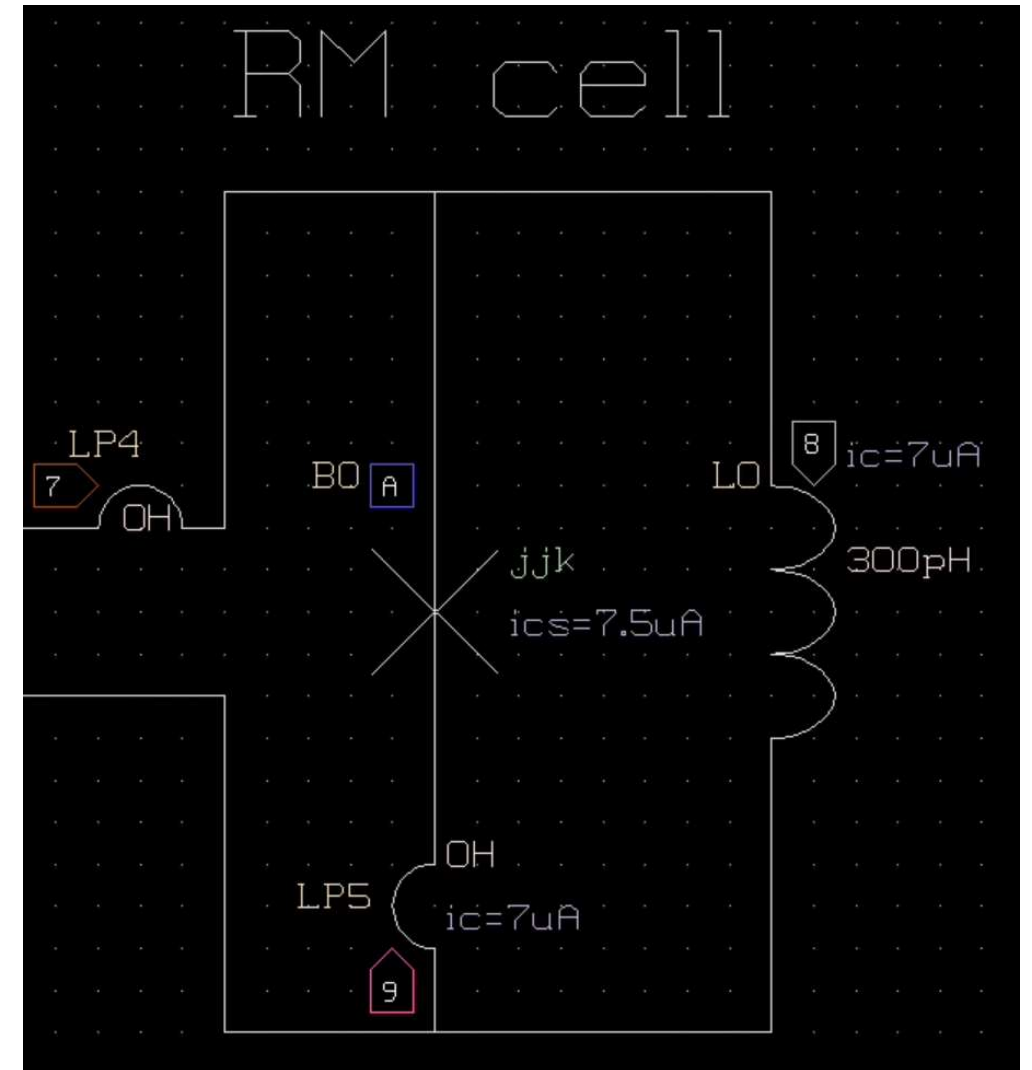
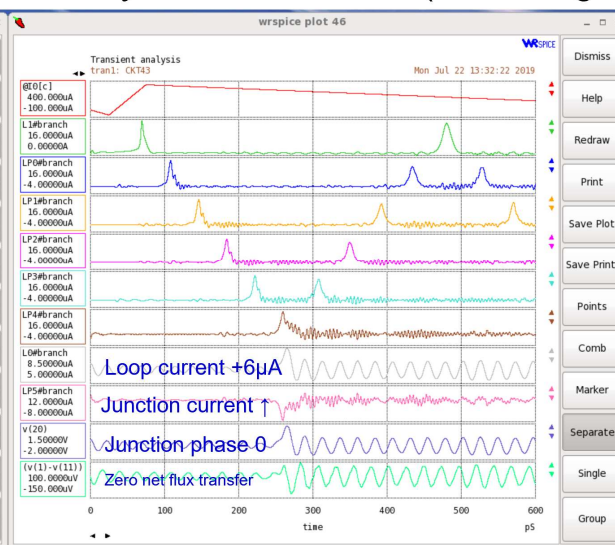
WRspice simulations with ± 1 fluxon initially in the loop

- Uses `ic` parameter, & `uic` option to `.tran` command
 - Produces initial ringing due to overly-constricted initial flux
 - Can damp w. small shunt G

Polarity mismatch \rightarrow Exchange



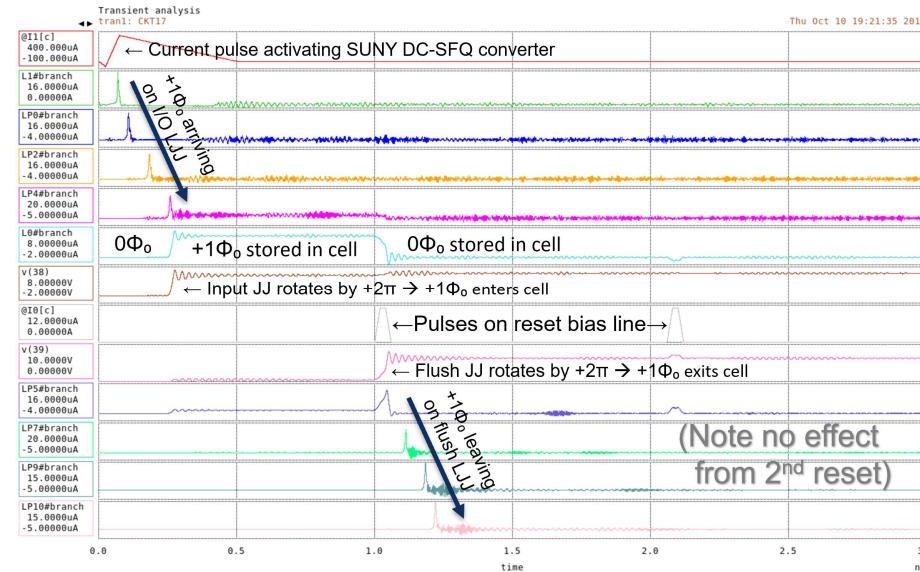
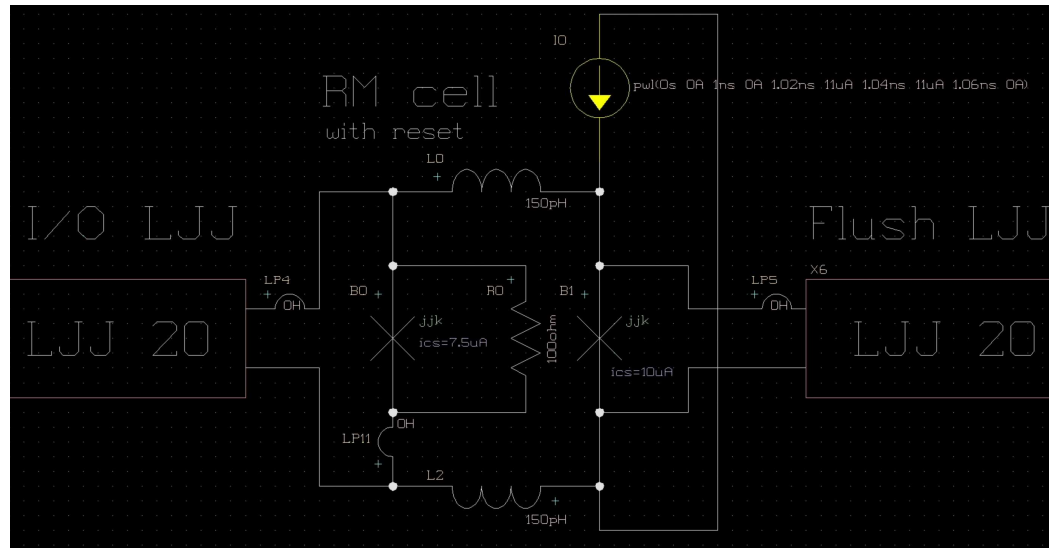
Polarity match \rightarrow Reflect (=Exchange)



Resettable version of RM cell—Designed & Fabricated!

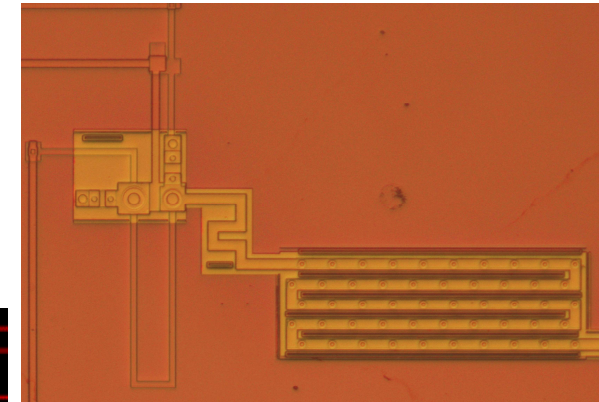
Apply current pulse of appropriate sign to flush the stored flux (the pulse here flushes out positive flux)

- To flush either polarity → Do both (\pm) resets in succession

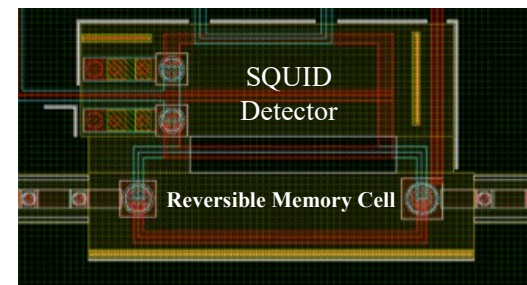
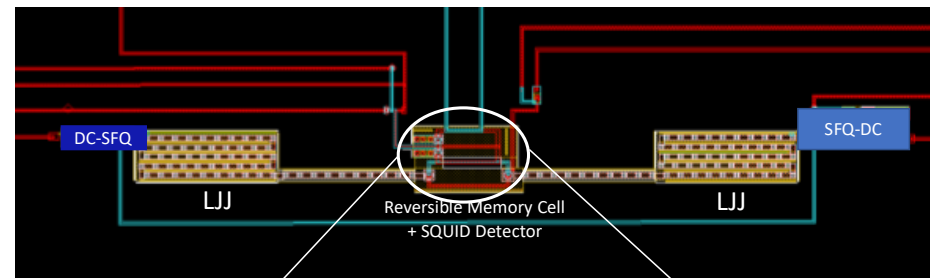
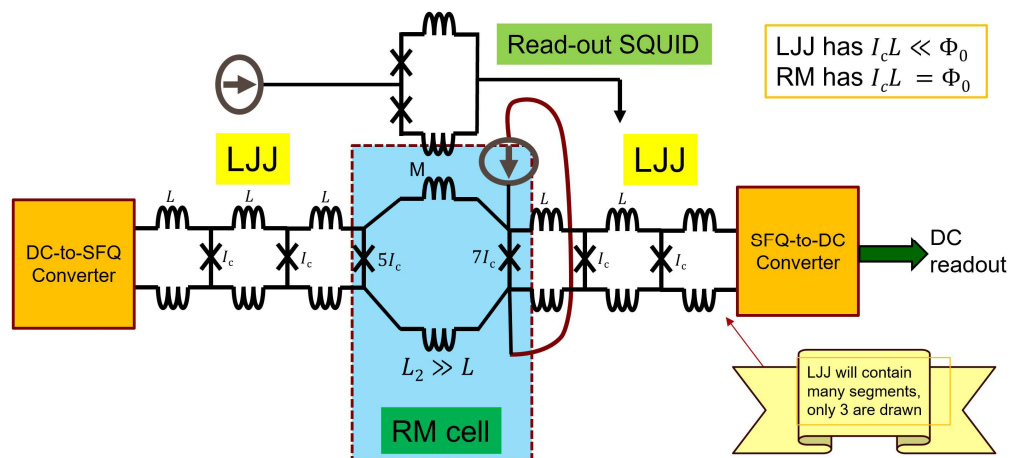
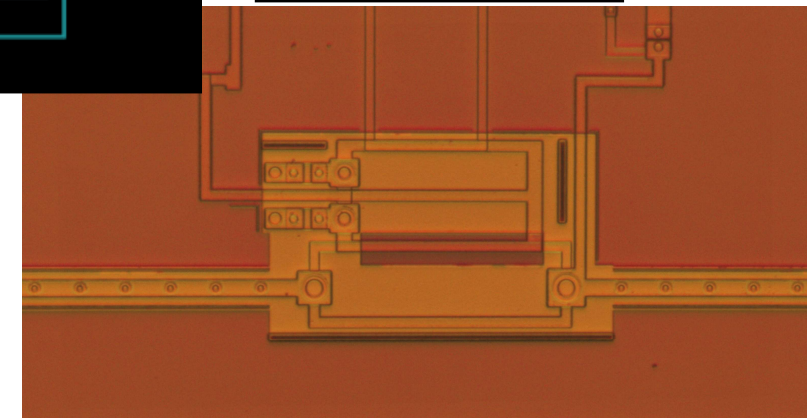


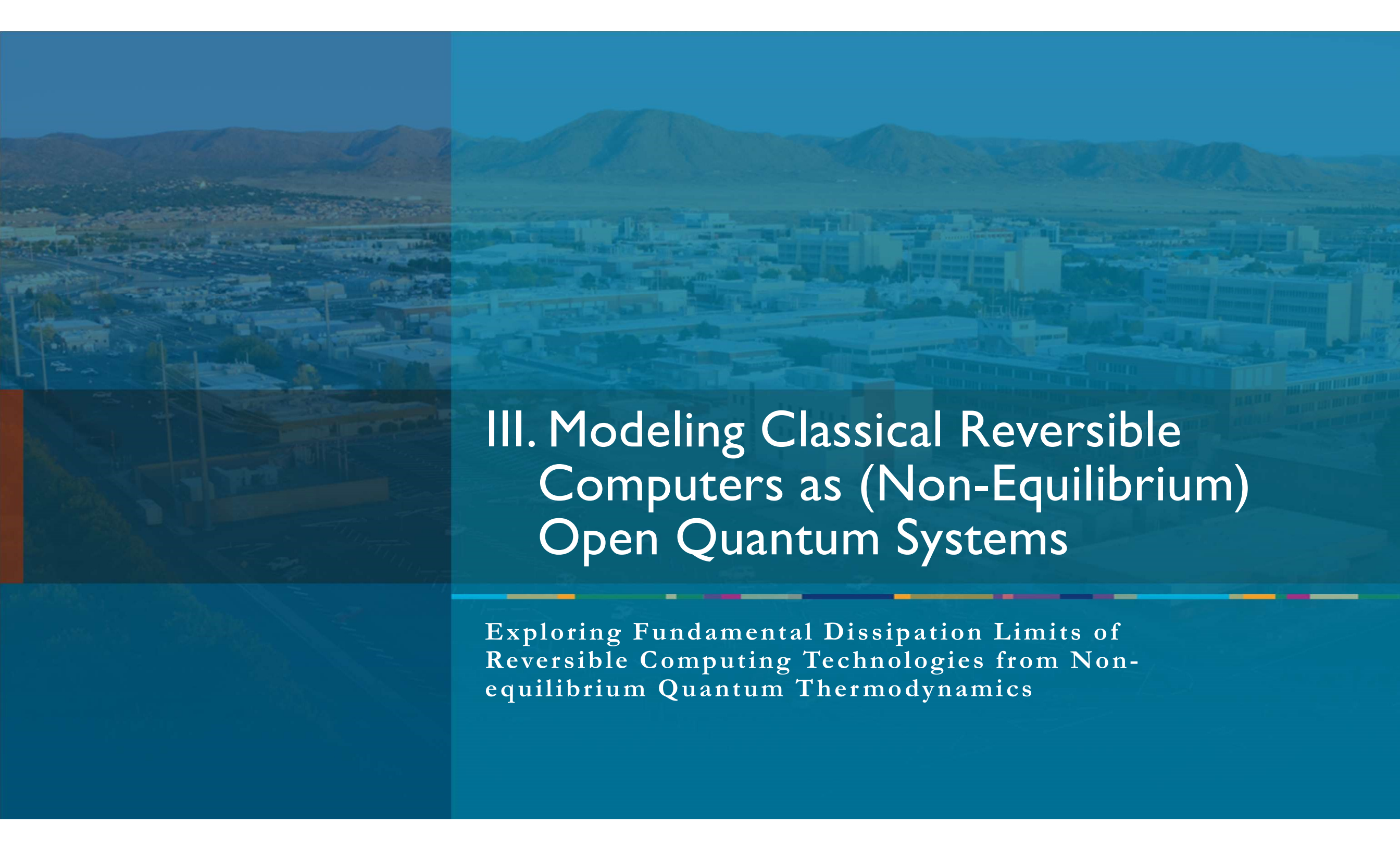
Fabrication at SeeQC with support from ACI

DC-SFQ & LJJ



RM Cell & SQUID





III. Modeling Classical Reversible Computers as (Non-Equilibrium) Open Quantum Systems

Exploring Fundamental Dissipation Limits of
Reversible Computing Technologies from Non-
equilibrium Quantum Thermodynamics

Modeling Classical Digital Computing Machines as Open Quantum Systems

From M. Frank & K. Shukla, *Entropy* 23(6), 701 (2021). [doi:10.3390/e23060701](https://doi.org/10.3390/e23060701)

(Cover article, special issue on “Physical Information and the Physical Foundations of Computation.”)

Basic open quantum systems picture:

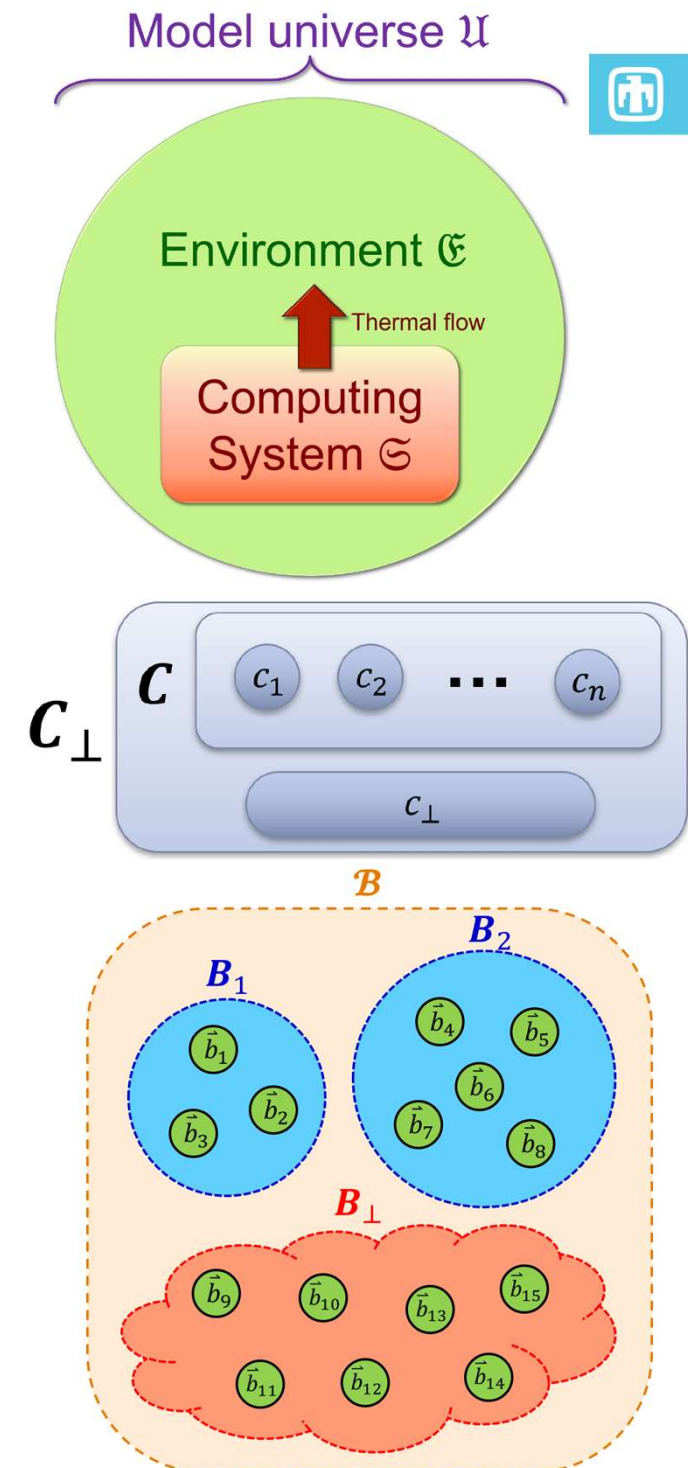
- Computer system \mathcal{S} with internal power supply, expels waste heat to thermal env. \mathcal{E} .
 - Idealize environment as very large, & as being in an equilibrium state, at some \sim constant temperature T .
 - As per Stinespring, explicit assumption that the environment does *not* preserve correlations (Markovian assumption).

Very simple, generic picture of the abstract *computational* state space:

- Some set $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$ of n distinct abstract computational states.
 - This set can be time-dependent, but might be defined only at discrete times; $\mathcal{C}(\tau_\ell)$ for integer index ℓ .
- Augmented with an extra state c_\perp meaning “the machine is not in any valid state.”

Important concept of a *proto-computational basis* for the system \mathcal{S} :

- This can be *any* o.n. basis \mathcal{B} for the Hilbert space $\mathcal{H}_{\mathcal{S}}$ of \mathcal{S} whose set of basis vectors $\{\vec{b}\}$ *partitions* into equivalence classes $\{\mathcal{B}_i\}$ for $i \in \{\perp, 1, 2, \dots, n\}$ s.t. for any $\vec{b} \in \mathcal{B}_i$, a quantum state $|\phi\rangle$ with $|\langle \vec{b} | \phi \rangle| = 1$ is *unambiguously* interpretable as representing c_i .
 - We can even have a time-dependent $\mathcal{B}(t)$, if it's convenient for the state representation to be changing in time.
- Note that for any given equivalence class \mathcal{B}_i , arbitrary superpositions of basis vectors in the basis subset \mathcal{B}_i *also* unambiguously represent c_i .
 - Each basis subset \mathcal{B}_i thus spans a subspace \mathcal{H}_i of $\mathcal{H}_{\mathcal{S}}$ corresponding to computational state $c = c_i$.
 - The subspace basis \mathcal{B}_i may be chosen to be *any* basis for \mathcal{H}_i (makes no difference).



Computational vs. Non-computational Subsystems of a Computer



Conceptually, we can divide up the computer system \mathcal{S} into two subsystems:

- The **computational subsystem** \mathcal{C} is an abstract subsystem holding the digital state, c .
- The **non-computational subsystem** \mathcal{N} carries *everything else* that makes up the complete physical state of system \mathcal{S} .

In general, the Hilbert space of \mathcal{N} *depends* on the computational state of \mathcal{C} .

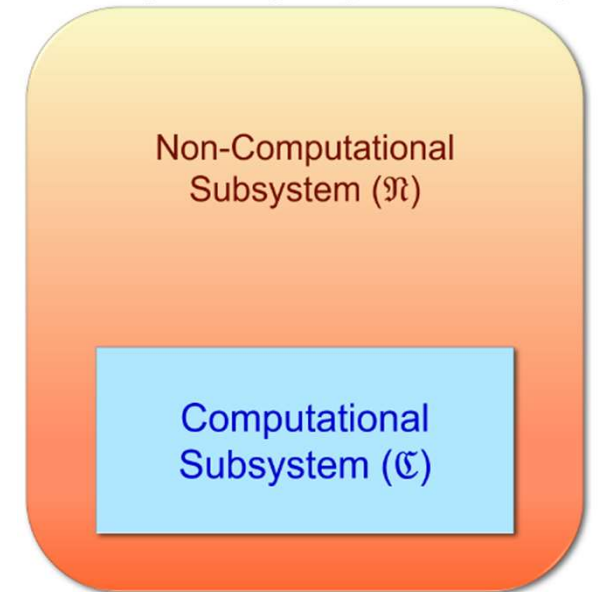
- The non-computational Hilbert space $\mathcal{H}_{\mathcal{N}}^c$ corresponding to computational state $c = c_i$ is simply that \mathcal{H}_i that's spanned by the protocomputational basis subset $B_i \subset \mathcal{B}$.

Then, formally speaking, the Hilbert space $\mathcal{H}_{\mathcal{S}}$ of the entire system \mathcal{S} can be expressed as a *subspace sum* of the non-computational Hilbert spaces $\mathcal{H}_{\mathcal{N}}^c$.

$$\mathcal{H}_{\mathcal{S}} = \bigoplus_{c \in \mathcal{C}_{\perp}} \mathcal{H}_{\mathcal{N}}^c$$

This simply means, a general vector in $\mathcal{H}_{\mathcal{S}}$ is a sum of vectors in the $\mathcal{H}_{\mathcal{N}}^c$, where the sub-bases for all the $\mathcal{H}_{\mathcal{N}}^c$ are all mutually orthogonal (as is true in our case).

Computing System (\mathcal{S})



$$\begin{aligned} \mathcal{C}_{\perp} &= \{c_1, \dots, c_n\} \\ B_i &= \{\vec{b}_{i,1}, \dots, \vec{b}_{i,n_i}\} \\ \vec{v}_i^{\mathcal{N}} &= \sum_{j=1}^{n_i} a_{i,j} \vec{b}_{i,j} \\ \vec{v}^{\mathcal{S}} &= \sum_{i=1}^n \sum_{j=1}^{n_i} a_{i,j} \vec{b}_{i,j} \end{aligned}$$

Representing the Quantum State of a Classical Computer as a Block-Diagonal Density Matrix



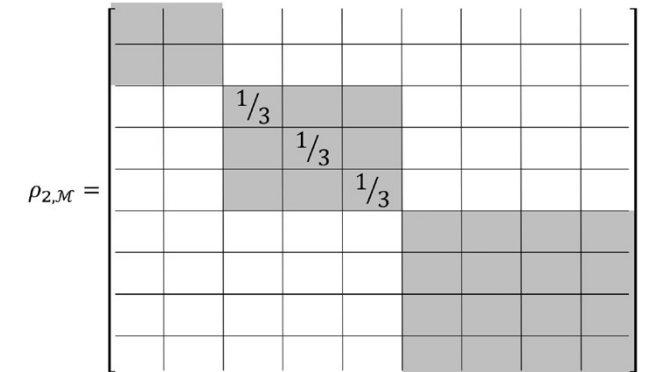
Each block in the density matrix represents a mixed quantum state of the non-computational subsystem \mathfrak{N}_c when the computational state is some particular $c = c_i$, weighted by the probability of that state.

We assume there are no coherences *between* blocks (i.e., decoherence is very fast).

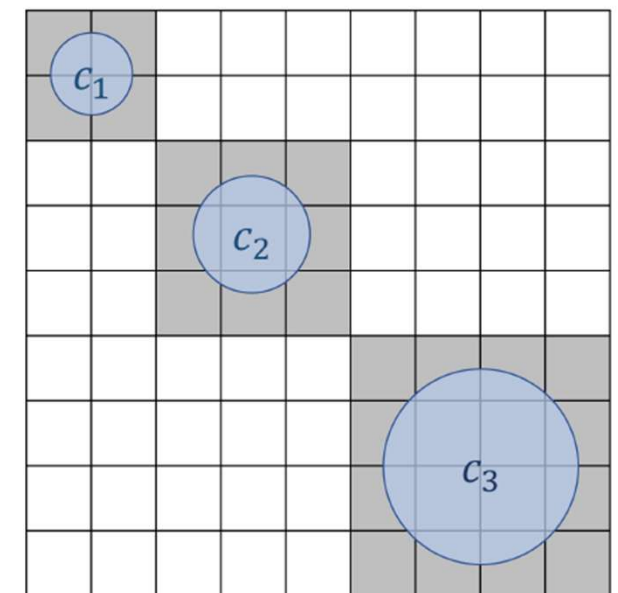
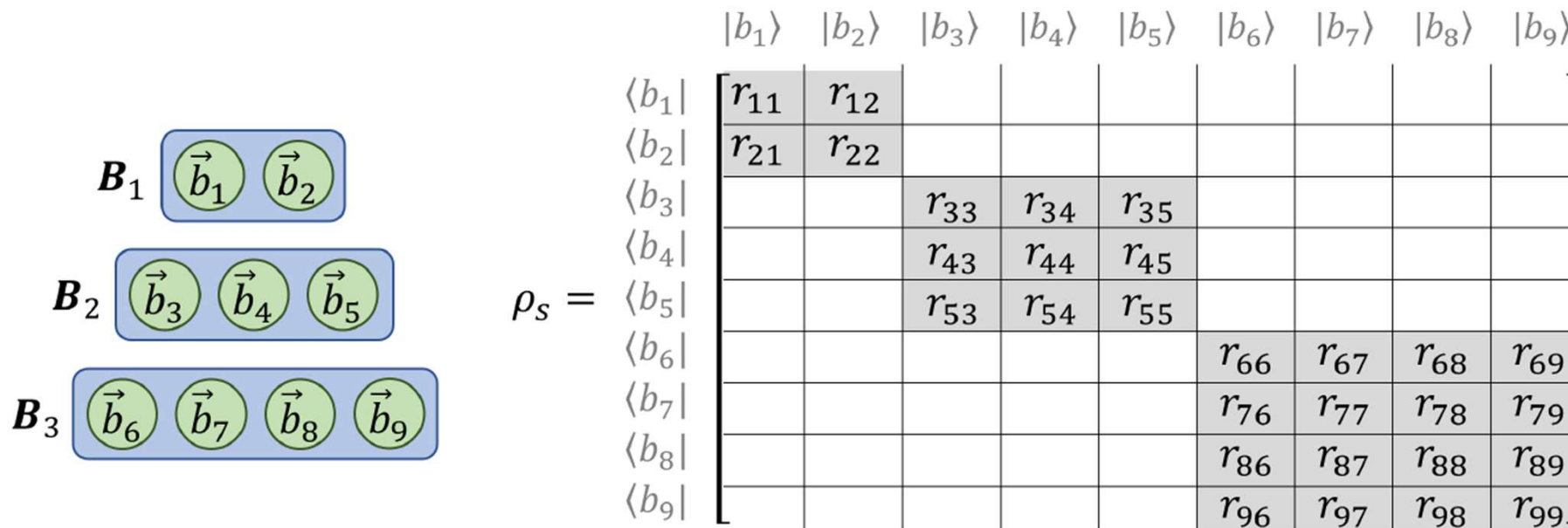
- That's why this is a classical computer, and not a quantum one!

Given that the computational states themselves are stable vs. decoherence, it follows that the blocks must contain the natural *pointer states* a.k.a. any *decoherence-free subspaces* of the system.

- Decoherence-free e.g. if they're *already* fully decohered by environment interaction.



Max-ent density matrix for state c_2



Computational Operations, Types of Operations, and Their Physical Representations



A classical **computational operation** $O = O_s^t$, in general, is just a (possibly partial) stochastic map between the computational state sets $\mathcal{C}(s)$ and $\mathcal{C}(t)$ at two discrete points in time $s = \tau_{k_1}$ and $t = \tau_{k_2}$; that is:

$$O_s^t: \mathcal{C}(s) \rightarrow \mathcal{P}[\mathcal{C}(t)],$$

where $\mathcal{P}[\cdot]$ denotes the normalized probability distributions over the given set.

Deterministic operations yield only single-valued (point) distributions.

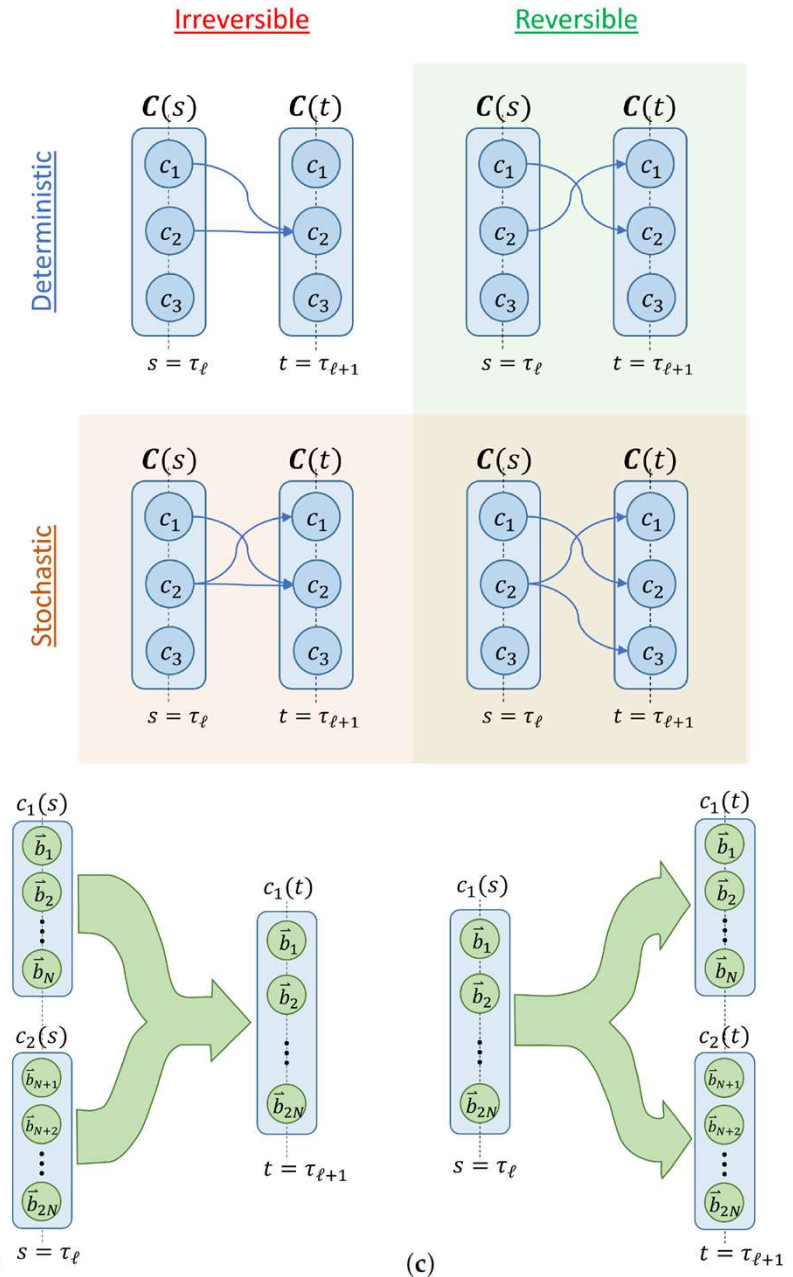
- whereas **stochastic** ones have at least one case with a multi-point distribution.

Reversible operations have all mutually non-overlapping distributions.

- while **irreversible** ones have at least one output state that's reachable with nonzero probability from at least two initial states in the allowed domain.

Physically speaking, computational operations are implemented by transformations of the basis subsets $B_i(\tau_k)$ corresponding to computational states $c_i(\tau_k)$.

- **Irreversible** operations imply *merging* of basis subsets.
- **Stochastic** operations imply *splitting* of basis subsets.



The Fundamental Theorem of the Thermodynamics of Computing



This theorem (FTTC) asserts the *inter-convertibility* of entropy between the computational and non-computational subsystems.

- By **merging** states, **irreversible** computational operations can move entropy from the computational subsystem to the non-computational subsystem.
- By **splitting** states, **stochastic** computational operations can move entropy from the non-computational subsystem to the computational subsystem.
- **NOTE:** We distinguish FTTC from Landauer's Principle proper!
 - L.P. refers more specifically to an information *loss* requiring an *increase* in total entropy.

Setup: Let $\phi \in \mathcal{H}_{\mathfrak{C}}$ represent a microstate (pure quantum state) of the computing system \mathfrak{C} .

- More precisely, let ϕ be hypothetically sampled by applying a complete projective measurement of \mathfrak{C} onto some protocomputational basis \mathcal{B} .
 - Thus ϕ can be identified as ϕ_i , corresponding to some $\vec{b}_i \in \mathcal{B}$.
- Probability distribution $p(\phi_i)$ is given by Born rule, or (equivalently) by the diagonal elements of the $\rho_{\mathfrak{C}}$ density matrix in the \mathcal{B} basis.

This implies a *derived* prob. distribution over the computational states:

$$P(\mathbf{c}_j) = \sum_{\phi_i \in \mathbf{c}_j} p(\phi_i).$$

And the total entropy of the physical state of the computer system (random variable Φ) can always be written as $S(\Phi) = H(C) + S(\Phi|C)$,

- where C is a random variable for the computational state, and S, H are the entropies based on the probability distributions p, P respectively.

Computing System (\mathfrak{C}),
total entropy $S(\Phi) = -\sum p \log p$

Non-Computational
Subsystem (\mathfrak{N})

non-computational /
conditional entropy
 $S_{\mathfrak{N}} = S(\Phi|C) = S(\Phi) - H(C)$

Computational
Subsystem (\mathfrak{C})

info. entropy $H(C) = -\sum P \log P$

Total entropy $S(\Phi)$ is always the sum of the computational entropy $H(C)$ and non-computational entropy $S(\Phi|C)$.

Proof of Landauer's Principle (example for correlated-subsystems case)

Let X, Y be state variables corresponding to *any two* disjoint computational subsystems $\mathfrak{X}, \mathfrak{Y}$ within a larger computer \mathfrak{C} .

- There is a joint probability distribution $P(X, Y)$, and a corresponding joint entropy $H(X, Y)$.
- Reduced (marginal) entropies $H(X), H(Y)$ of the individual subsystems are defined in the usual way. (Note, this are not, in general, “true” entropy!)

The mutual information between \mathfrak{X} and \mathfrak{Y} is defined as:

$$I(X; Y) \stackrel{\text{def}}{=} H(X) + H(Y) - H(X, Y).$$

Now, define the “*independent entropy in \mathfrak{Y}* ” as the *rest* of \mathfrak{Y} 's (reduced subsystem) entropy, *besides* the mutual information I that \mathfrak{Y} has with \mathfrak{X} :

$$S_{\text{ind}}(Y) \stackrel{\text{def}}{=} H(Y) - I(X; Y) = H(Y|X),$$

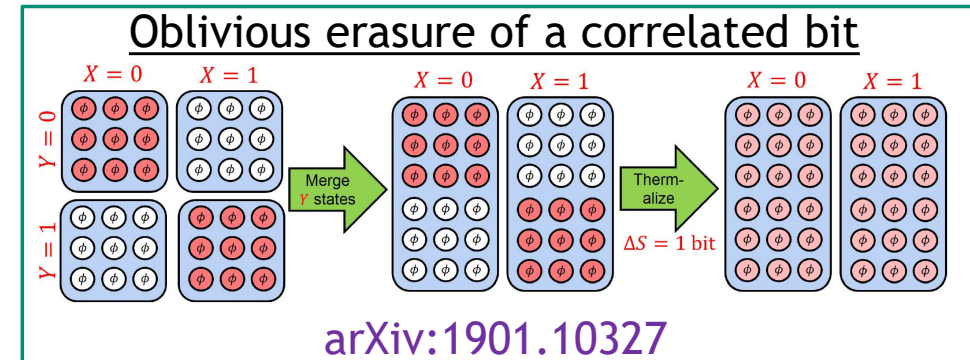
- This is just the same as the conditional entropy of \mathfrak{Y} , conditioned on \mathfrak{X} .

Now, consider *erasing \mathfrak{Y}* via *any oblivious* physical mechanism...

- Meaning, force $H(Y) = 0$ *unconditionally*, without making use of X *or any other information we may have* about Y .
 - E.g., remove a potential energy barrier separating $Y = 0$ and $Y = 1$ computational states, and interpret the new merged computational state as meaning $Y = 0$.

And assume, in general, *non-computational information* (in \mathfrak{N}) *will fairly rapidly thermalize*. (If not, then why even consider it non-computational?)

- This thermalization process (which occurs by the time information is ejected to \mathfrak{C}) is when/where the absolute entropy increase happens in Landauer!
 - By assumption, environment evolution is not tracked, ergo any \mathfrak{C} - \mathfrak{C} correlation is lost.



Note that we could try to “reverse” the whole erasure process to restore the original reduced entropy $H(Y)$ of the \mathfrak{Y} subsystem...

But now, $I(X; Y)_{\text{new}} = 0$ (any correlations have become lost!)

- $\therefore S_{\text{ind}}(Y) = H(Y)$, $\therefore \Delta S_{\text{ind}}(Y) = I(X; Y)_{\text{orig}} = \Delta S_{\text{tot}}$.

If, originally, Y was (*deterministically*) computed from X , then:

- $H(Y|X)_{\text{orig}} = 0$, i.e., $S_{\text{ind}}(Y) = 0$, so $H(Y) = I(X; Y)_{\text{orig}}$.
 - *Apparent* entropy of *all* computed bits is *actually entirely* mutual information!
 - a.k.a. “information-bearing entropy” in Anderson’s terminology

Independent entropy (and total universe \mathfrak{U} entropy!) increased by

$$\Delta S_{\text{tot}} = \Delta S_{\text{ind}}(Y) = I(X; Y)_{\text{orig}} = H(Y).$$

\therefore Erasing computed (as opposed to *random*) bits in isolation (without using knowledge/correlations) turns their *digital* information into new physical entropy.

Q.E.D.! ■

What does it mean for the unitary time-evolution of a system to implement a classical computation?



Given our framework, this has a very natural, straightforward definition.

- This is more formally defined in our paper, but below is a quick informal description.

We say that the quantum time-evolution of a given computer system \mathcal{G} *correctly implements* the classical computational operation O between times s and t for the initial mixed state ρ_s if and only if, after the unitary time evolution U_s^t has occurred, and we adjust as needed for any time-dependence in the protocomputational basis \mathcal{B} , and we allow the resulting state to decohere naturally, we end up with the correct probability distribution $P(c(t))$ over the final computational states $c(t)$ that's implied by applying O_s^t to the initial-state distribution $P(c(s))$ that's implied by ρ_s .

- In particular, we don't care about *any* details of the resulting quantum state ρ_t *other* than the overall distribution P over classical computational states that it implies.

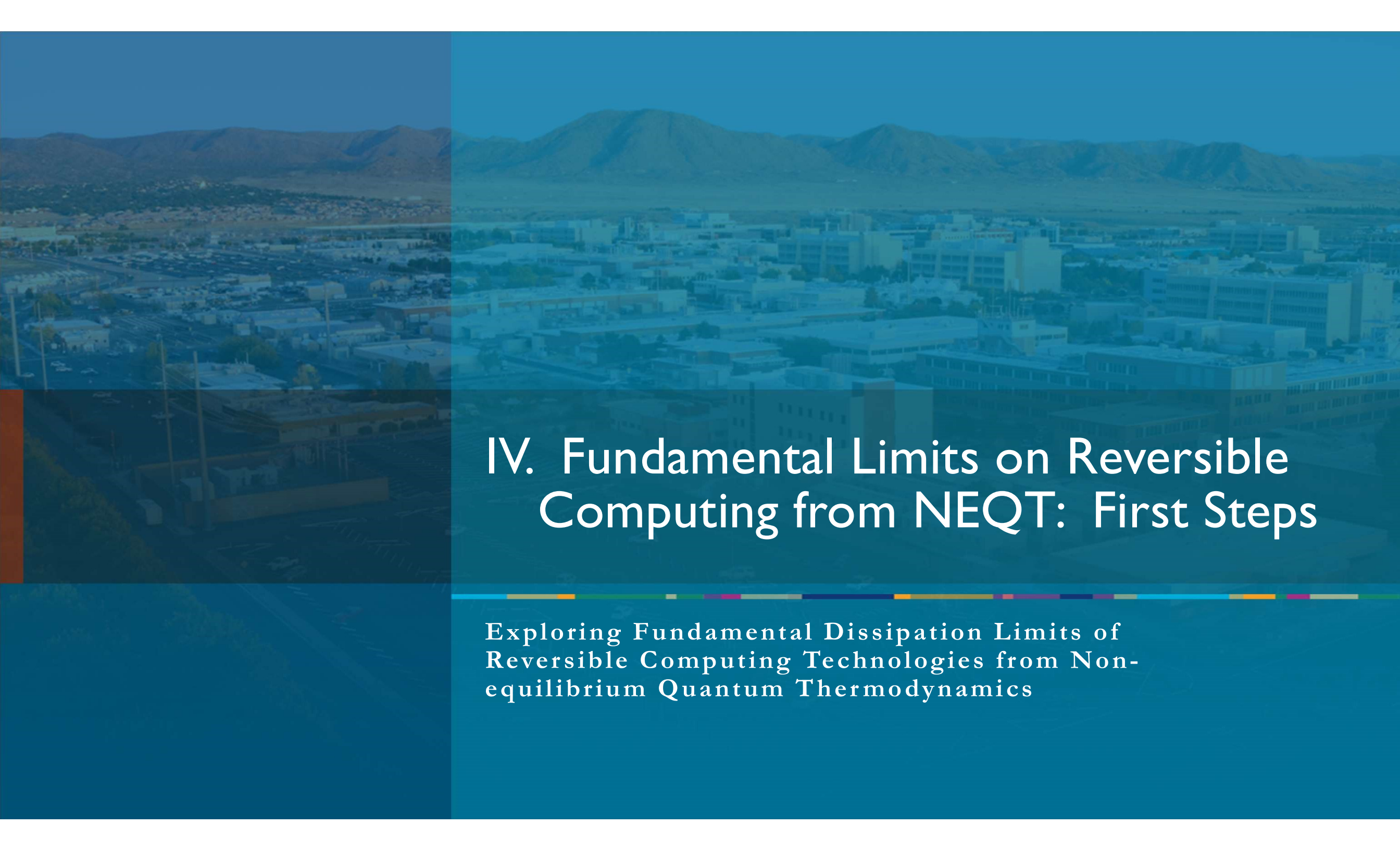
Here's a concise notation for expressing the above condition:

$$\underbrace{U_s^t(\mathcal{G}, \mathcal{B})}_{\text{Unitary time evolution of computer system } \mathcal{G} \text{ from time } s \text{ to } t, \text{ with a possible change in protocomputational basis } \mathcal{B}(s) \rightarrow \mathcal{B}(t).} \Vdash \underbrace{\mathcal{C}_s^t(O_s^t, \rho_s)}_{\text{A computation, defined as the abstract operation } O: \mathcal{C}(s) \rightarrow \mathcal{C}(t) \text{ being performed from the starting quantum statistical operating context } \rho_s.}$$

Unitary time evolution of computer system \mathcal{G} from time s to t , with a possible change in protocomputational basis $\mathcal{B}(s) \rightarrow \mathcal{B}(t)$.

A computation, defined as the abstract operation $O: \mathcal{C}(s) \rightarrow \mathcal{C}(t)$ being performed from the starting quantum statistical operating context ρ_s .

“implements”



IV. Fundamental Limits on Reversible Computing from NEQT: First Steps

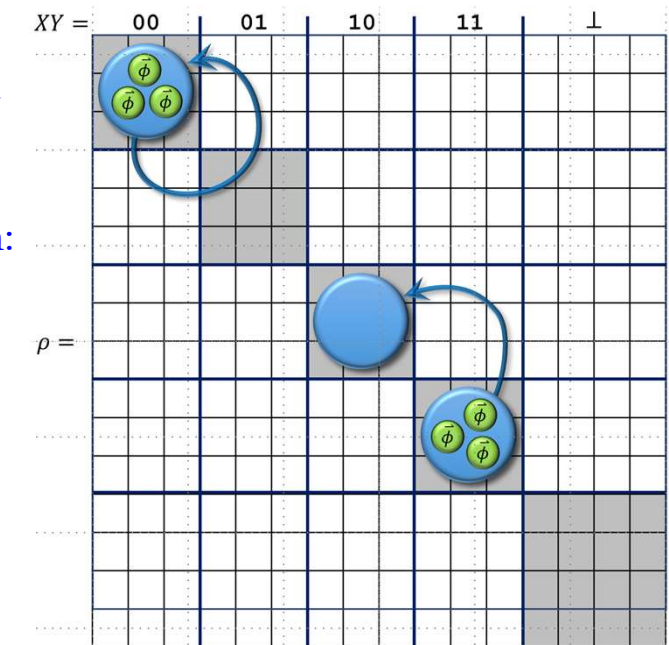
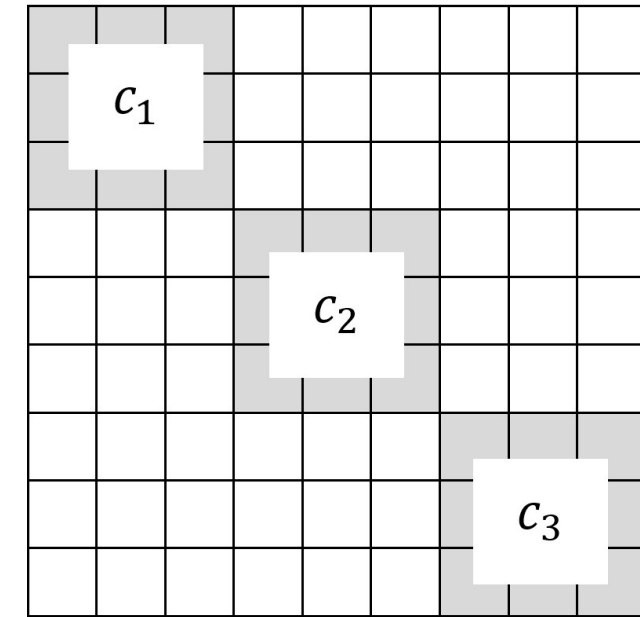
Exploring Fundamental Dissipation Limits of Reversible Computing Technologies from Non-equilibrium Quantum Thermodynamics

Reversible Operations as Quantum Channels

(Work by K. Shukla,
Brown U.
doi:10.3390/e23060701)

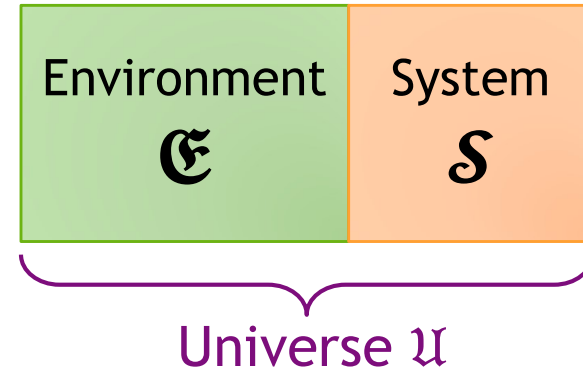


- Want to characterize dissipation of reversible operations.
 - Do fundamental limits exist? If so, what is the dependence on fundamental parameters?
 - *Most general* limits for practical models: Use nonequilibrium quantum thermodynamics (NEQT).
 - Unitary evolution: No dissipation, but operation time is bounded by quantum speed limit (QSL).
 - Dissipation as a function of delay ($D(d)$). Goal: retrieve protocol-based, device-independent expression.
 - Note that, in principle, $D(d)$ is not directly determined by QSL (since not all energy invested need be dissipated), but we can use QSL to obtain a *non-tight* upper bound on the required dissipation.
- For quantum limits: Natural framework: Represent classical operations as quantum channels.
 - Computational states c_i form equivalence classes over (physical) quantum states $|\psi\rangle$. Coherences allowed between different $|\psi\rangle$ corresponding to the same c_i , but not between $|\psi\rangle$ corresponding to different c_i .
 - Thus, each c_i is a single decoherence-free subspace (DFS) of overall Hilbert space. Classical computation: operations transforming states from one block to another, with no states leaving the block structure.
 - Computation embedded in open system. Information can “leak” into environment, but (we assume) cannot be recaptured at any future time. Thus, dynamics represented by Markovian (GKSL) evolution.
 - GKSL dynamics with multiple asymptotic states (V. V. Albert et al. Phys. Rev. X 6, 041031 (2016)): asymptotic states form subspace in overall dynamics. Provides most general embedding of quantum channels in GKSL structure.



Conditional Landauer Reset of Y / De-computing Y from X (Reversible)

Stinespring Dilation Theorem and Thermal Operations



- Stinespring dilation theorem: provides a representation of quantum channels, by embedding them in a larger space.
- If $\rho \in \mathcal{D}(\mathcal{H}_{\mathcal{S}})$ is the state of system \mathcal{S} , can represent any transformation $\rho \mapsto \Lambda_t[\rho]$ by examining joint unitary evolution on any larger system that contains $\mathcal{S}\mathfrak{E}$.
 - Let $\mathfrak{U} = \mathcal{S}\mathfrak{E}$ be universe, comprising \mathcal{S} in initial state $\rho_{\text{in},\mathcal{S}}$ and env. \mathfrak{E} in initial state $\rho_{\text{in},\mathfrak{E}}$.
 - Time evolution of $\rho_{\text{in},\mathcal{S}}$ is given by:

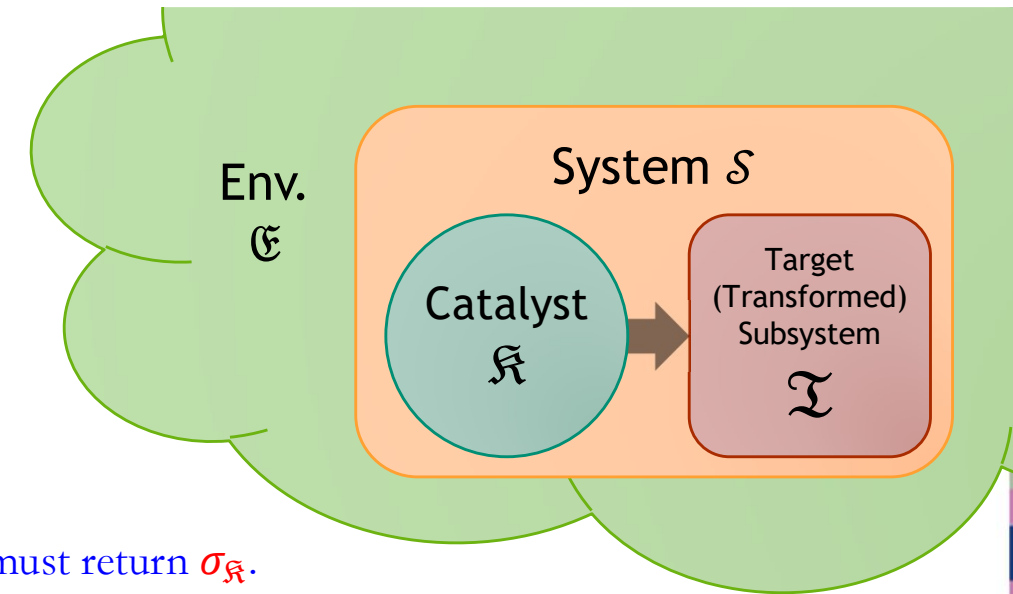
$$\rho_{\text{in},\mathcal{S}} \mapsto \Lambda_t[\rho_{\text{in},\mathcal{S}}] := \text{Tr}_{\mathfrak{E}}[\hat{U}_{t,\mathfrak{U}}(\rho_{\text{in},\mathcal{S}} \otimes \rho_{\text{in},\mathfrak{E}})\hat{U}_{t,\mathfrak{U}}^\dagger] = \text{Tr}_{\mathfrak{E}}[e^{-i\hat{H}_{\mathfrak{U}}(t_f - t_0)}(\rho_{\text{in},\mathcal{S}} \otimes \rho_{\text{in},\mathfrak{E}})e^{i\hat{H}_{\mathfrak{U}}(t_f - t_0)}]$$

- Note, final trace here is over \mathfrak{E} . Thus, map $\Lambda_t[\rho_{\text{in},\mathcal{S}}] \in \mathcal{D}(\mathcal{H}_{\mathcal{S}})$ is also a density matrix over \mathcal{S} .
- Thermal operations (TOs): set of all (thermodynamically) possible transformations on ρ that can be implemented at no energetic cost. Given by setting initial state of \mathfrak{E} as thermal (Gibbs) state $\tau_{\mathfrak{E}}$:

$$\rho_{\text{in},\mathcal{S}} \mapsto \Xi_t[\rho_{\text{in},\mathcal{S}}] := \text{Tr}_{\mathfrak{E}}[\hat{U}_{t,\mathfrak{U}}(\rho_{\text{in},\mathcal{S}} \otimes \tau_{\mathfrak{E}})\hat{U}_{t,\mathfrak{U}}^\dagger]$$

- Necessary conditions for TOs and for catalytic TOs (next slide) described by resource theory of quantum thermodynamics (RTQT).
 - Provides free states (*i.e.*, states that can be generated at no thermodynamic cost) and free operations (TOs).
 - References for introduction to RTQT: N. H. Y. Ng and M. P. Woods, in *Thermodynamics in the Quantum Regime*, ed. by F. Binder *et al.* (Springer Nature, Cham, 2018) and M. Lostaglio, Rep. Prog. Phys. **82**, 114001 (2019).

General Catalytic Thermal Operations



- To model transformations involving a *catalyst*, we can extend the notion of TOs to **catalytic thermal operations** (CTOs).
- Divide system \mathcal{S} into subsystems \mathcal{I} and \mathcal{K} .
- Catalyst \mathcal{K} is required for transformation $\rho_{\text{in}, \mathcal{I}} \mapsto \mathbb{E}_t[\rho_{\text{in}, \mathcal{I}}]$ on \mathcal{I} .
- Catalyst locally starts & ends in the same state:
 - If $\sigma_{\mathcal{K}}$ is initial state of \mathcal{K} , then partial trace over $\mathcal{I}\mathcal{E}$ after global unitary evolution must return $\sigma_{\mathcal{K}}$.
- Most general type of CTO (M. Müller, Phys. Rev. X **8**, 041051 (2016)): Start with $\rho_{\text{in}, \mathcal{I}}$ as initial state of \mathcal{I} and $\sigma_{\mathcal{K}}$ as initial state of \mathcal{K} (with $\rho_{\text{in}, \mathcal{I}} \otimes \sigma_{\mathcal{K}}$ as initial state of $\mathcal{S} = \mathcal{I}\mathcal{K}$), and with thermal state $\tau_{\mathcal{E}}$ of environment \mathcal{E} as initial state of \mathcal{E} . Most general CTO on \mathcal{S} are given by:

$$(\rho_{\text{in}, \mathcal{I}} \otimes \sigma_{\mathcal{K}}) \mapsto \xi_{\mathcal{I}\mathcal{K}} := \text{Tr}_{\mathcal{E}}[\hat{U}_{t, \mathcal{I}\mathcal{K}\mathcal{E}}(\rho_{\text{in}, \mathcal{I}} \otimes \sigma_{\mathcal{K}} \otimes \tau_{\mathcal{E}})\hat{U}_{t, \mathcal{I}\mathcal{K}\mathcal{E}}^\dagger].$$

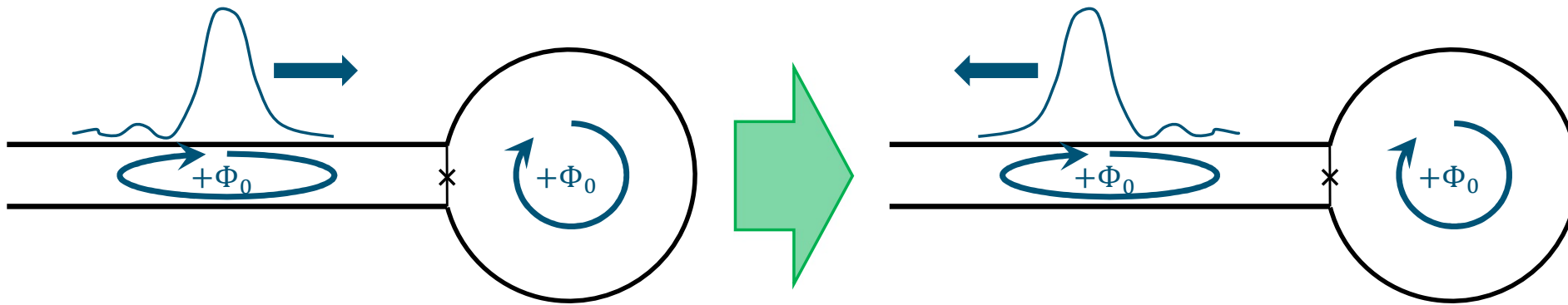
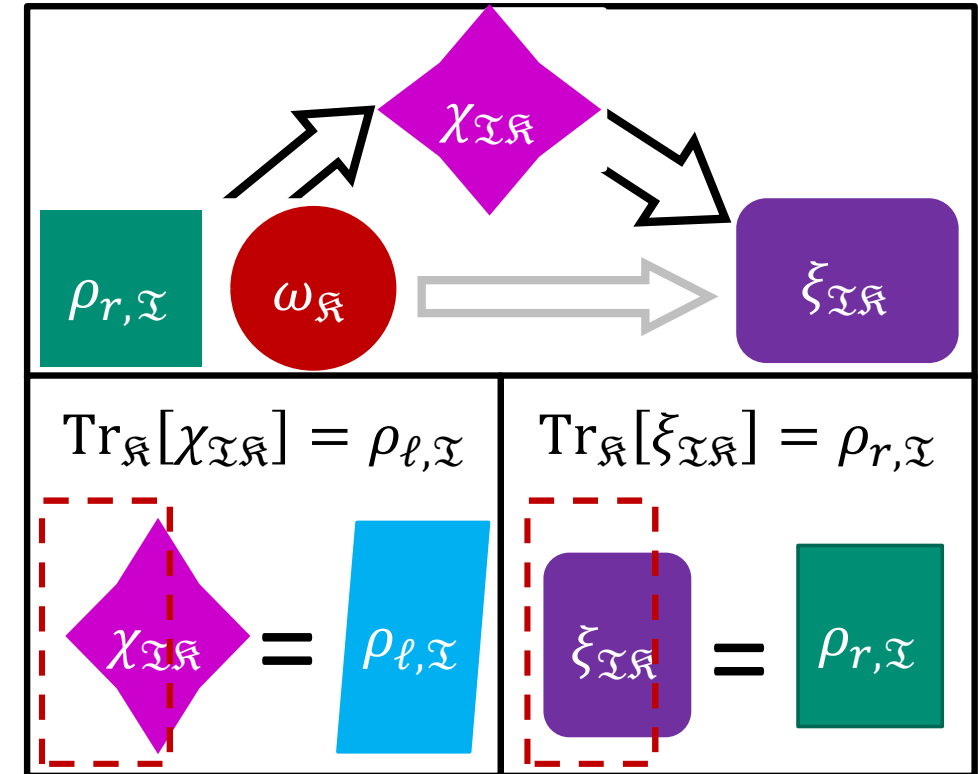
- $\|\text{Tr}_{\mathcal{I}\mathcal{E}}[\xi_{\mathcal{I}\mathcal{K}}] - \mathbb{E}_t[\rho_{\text{in}, \mathcal{I}}]\|_1 < \epsilon$ for any $\epsilon \in \mathbb{R}^+$ (CTO corresponds to desired transformation $\rho_{\text{in}, \mathcal{I}} \mapsto \mathbb{E}_t[\rho_{\text{in}, \mathcal{I}}]$ locally on \mathcal{I})
- $\text{Tr}_{\mathcal{I}\mathcal{E}}[\xi_{\mathcal{I}\mathcal{K}}] = \sigma_{\mathcal{K}}$ (catalyst must locally end in the same state it started) and $[\hat{U}_{t, \mathcal{I}\mathcal{K}\mathcal{E}}, \hat{H}_{\mathcal{I}\mathcal{K}\mathcal{E}}] = 0$.
- Realizable if and only if Helmholtz free energy decreases; i.e., if $\mathcal{F}(\text{Tr}_{\mathcal{I}\mathcal{E}}[\xi_{\mathcal{I}\mathcal{K}}]) \leq \mathcal{F}(\rho_{\text{in}, \mathcal{I}})$ for $\mathcal{F}(\rho) := \text{Tr}_{\mathcal{I}\mathcal{E}}[\xi_{\mathcal{I}\mathcal{K}}]$.
- Quantum mutual info (QMI) between \mathcal{I} and \mathcal{K} after operation can be made as small as possible (but not zero):
For any $\delta \in \mathbb{R}^+$, there exists some \mathcal{K} and $\xi_{\mathcal{I}\mathcal{K}}$ such that:

$$S(\xi_{\mathcal{I}\mathcal{K}} \| \text{Tr}_{\mathcal{K}\mathcal{E}}[\xi_{\mathcal{I}\mathcal{K}}] \otimes \text{Tr}_{\mathcal{I}\mathcal{E}}[\xi_{\mathcal{I}\mathcal{K}}]) < \delta.$$
- Reversible computing can be modeled using this most general form of CTOs. (Preservation of correlation (QMI) between \mathcal{I} and \mathcal{K} avoids the Landauer cost that would be incurred if this QMI were ejected into the environment.)
- Also, **classical** information processing (IP) exhibits a lower bound on dissipation than the more general quantum case for IP operations (D. Bedingham and O. Maroney, New J. Phys. **18**, 113050 (2016)).

Computational Models from Generalized CTOs



- By composing two general CTOs, \mathcal{I} can start in a reset state $\rho_{r,\mathcal{I}}$, evolve to a new state $\rho_{\ell,\mathcal{I}}$ (corresponding to a computation), and then be reset by the catalyst back to the reset state. (See figure \rightarrow)
- In standard CTOs, we focus on a subsystem of interest and a catalyst.
 - However, for a system where multiple informational degrees of freedom propagate independently of external control mechanisms, each information carrier is simultaneously a subsystem of interest in its own right, and can act like a catalyst for the other carriers.
 - Ex.: ballistic reversible computing models.
 - Can further generalize CTOs to model computational processes in which interactions between N information carriers yield transformations on all N subsystems jointly.
- We can also generalize the notion of a catalyst with a single reset state to a subsystem with n distinguished states, transforming another system.
 - This then can model the computational notion of a *finite state machine*.
 - Example: In BARC, the circuit elements are Mealy machines transforming I/O symbols.



In ballistic asynchronous reversible computing (BARC), a part of the system that is unchanged by a given interaction can be considered a catalyst for transformation of other parts.

Applications to Quantum Computing

- Our theoretical framework is equally (and arguably, even more!) applicable to examining the thermodynamics of *quantum* computing:
 - Quantum channel embedding: Any quantum channel can be embedded in an appropriate asymptotic subspace. (V. Albert thesis, Sec. 2.1.4.)
 - Systems with a single (or multiple) DFS blocks can encode multiple qubit systems. (V. Albert thesis, Ch. 3.)
 - Thus, extending the thermodynamic dissipation length and TURs to multiple NESSs can allow us to calculate these quantities for *any* quantum channel we wish!
 - A quantum computation can be viewed as a quantum channel that transforms the states in transit.
- Further, even our main goal of developing more efficient technologies for *classical* computing can find eventual application in engineering more effective low-power digital systems for embedded cryogenic control of quantum architectures.

Can dissipation scale better than linearly with speed?



Some observations from Pidaparthi & Lent (2018) suggest Yes!

- Landau-Zener (1932) formula for quantum transitions in e.g. scattering processes with a missed level crossing...
 - Probability of exciting the high-energy state (which then decays dissipatively) scales down *exponentially* as a function of speed...

$$P_D = e^{-2\pi\Gamma}$$
 - This scaling is commonly seen in many quantum systems!
- Thus, dissipation-delay *product* may have *no lower bound* for quantum adiabatic transitions—if this kind of scaling can actually be realized in practice.
 - I.e.*, in the context of a complete engineered system.
- Question:** Will unmodeled details (e.g., in the driving system) fundamentally prevent this, or not?

J. Low Power Electron. Appl. 2018, 8(3), 30; <https://doi.org/10.3390/jlpea8030030>

Open Access Article

Exponentially Adiabatic Switching in Quantum-Dot Cellular Automata

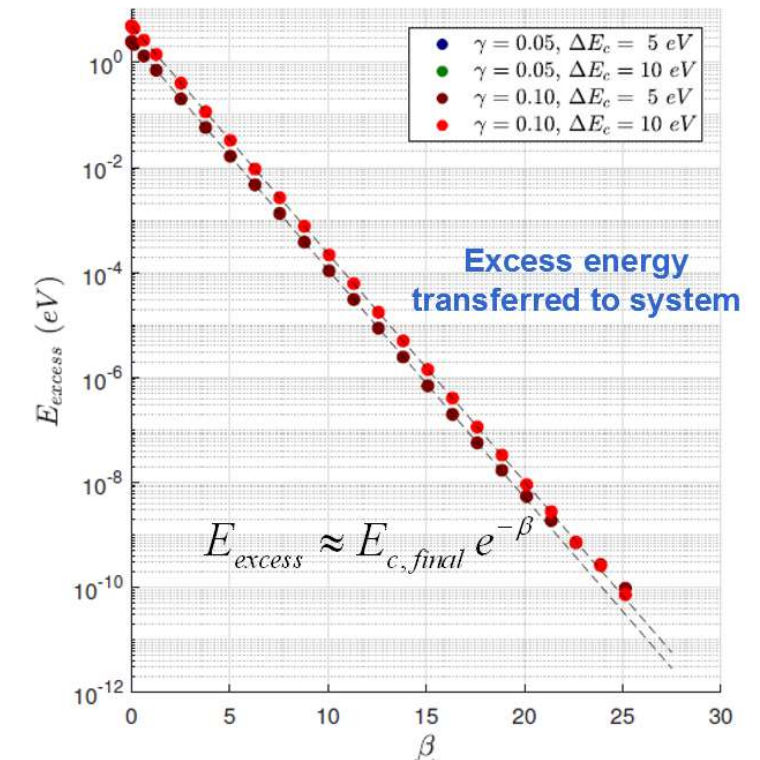
Subhash S. Pidaparthi and Craig S. Lent *

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

* Author to whom correspondence should be addressed.

Received: 15 August 2018 / Revised: 5 September 2018 / Accepted: 5 September 2018 / Published: 7 September 2018

(This article belongs to the Special Issue Quantum-Dot Cellular Automata (QCA) and Low Power Application)



The “Sweet Spot” in Operation Speed in Between the Relaxation and Thermal Coupling Timescales



Journal of
Applied Physics

Energy dissipation during two-state switching for quantum-dot cellular automata

Cite as: J. Appl. Phys. **129**, 024304 (2021); <https://doi.org/10.1063/5.0033633>

Submitted: 27 October 2020 • Accepted: 21 December 2020 • Published Online: 12 January 2021

Subhash S. Pidaparthi and Craig S. Lent

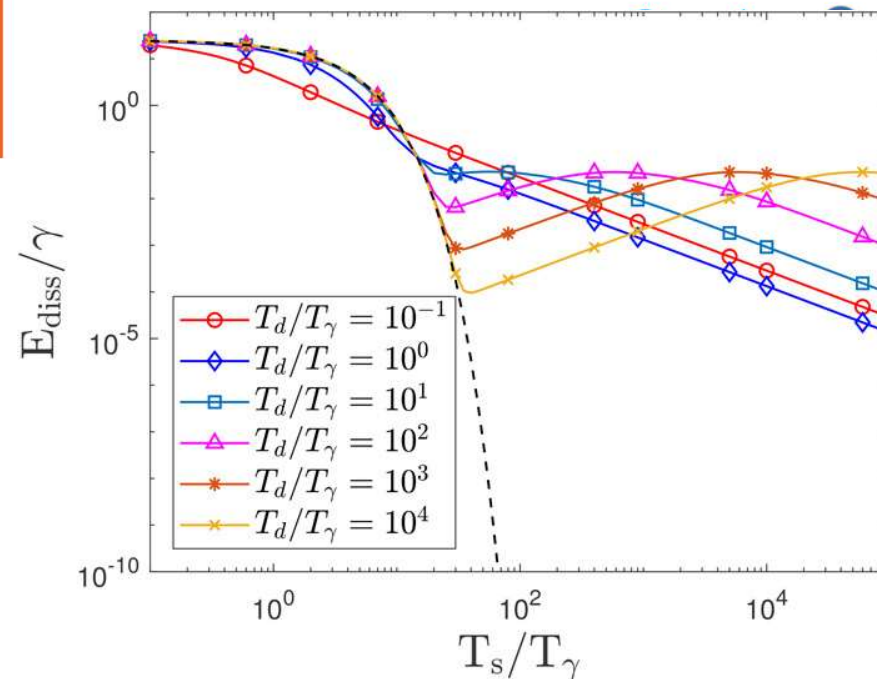


FIG. 10. Dissipated energy of an open system as a function of switching speed for different dissipation time constants. The dashed line is the excess energy of an isolated system. Here, the environmental temperature $k_B T / \gamma = 0.5$.

Published in: Subhash S. Pidaparthi; Craig S. Lent; *Journal of Applied Physics* **129**, 024304 (2021)
DOI: 10.1063/5.0033633
Copyright © 2021 Author(s). (Excerpted with permission.)

Lesson: When thermal coupling is low, dissipation from a dynamical state transition can be substantially suppressed compared to the background linear adiabatic scaling.

A similar result for ballistic superconducting circuits in Crutchfield, “Sub-Gigahertz Landauer Momentum Computing,” [arXiv:2202.07122](https://arxiv.org/abs/2202.07122):

Optimal τ^* s are upper bounded: the devices must operate *faster* than particular timescales—timescales determined by the substrate physics. The bit swap’s low work cost requires operating on a timescale faster than the rates at which the system exchanges energy and information with the environment. Thus, momentum computing protocols have a *speed floor* rather than a speed limit.

Conclusion

Reversible computing will be essential to maintain ongoing improvements in the energy efficiency of general digital computing.

- This follows from Landauer's principle; but *non*-reversible computing approaches may reach practical limits even before the efficiency bound from Landauer's limit is reached.

As we approach the limits of conventional computing, understanding the fundamental physical limits of reversible computing will become increasingly important in the coming decades.

- Essential if we want guidance in how to develop new physical mechanisms for computing that approach these limits as closely as possible.

The physics of reversible computing is a greatly under-studied topic that is ripe for increased attention.

- A wide array of powerful theoretical tools from fields such as non-equilibrium quantum thermodynamics (NEQT) are available to tackle this subject.

In this talk, we have surveyed the theoretical approach we are pursuing to better understand the quantum thermodynamic limits of reversible computing.

- Along the way, we reviewed why the classic insights of Landauer and Bennett can easily be seen to be completely valid in a modern theoretical framework.
- Much work remains to be done, and we invite interested collaborators to join us!

